Attorney's Docket No. 035718/237005 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.:	10/032,717	Confirmation No.: 5409
Applicant(s):	Abad *et al.*	
Filed:	October 23, 2001	
Art Unit:	1638	
Examiner:	Kubelik, Anne R.	
Title:	GENES ENCODING NOVEL BACILLUS THURINGIENSIS PROTEINS WITH PESTICIDAL ACTIVITY AGAINST COLEOPTERANS	
Docket No.:	035718/237005	
Customer No.:	29122	

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF UNDER 37 CFR § 41.37

This Appeal Brief is filed pursuant to the "Notice of Appeal to the Board of Patent Appeals and Interferences" filed August 11, 2006.

It is noted at the outset that this appeal is the third time that Applicants have attempted to appeal the decision of the Examiner in the present application. In the two previous attempts, the Examiner has reopened prosecution. In the most recent appeal, prosecution was reopened after the filing of a Reply Brief on February 23, 2006, and the rejections by the Examiner are identical to those addressed in the Appeal Brief of September 6, 2005. The Examiner cited additional references to support her arguments. However, the references have been available since before the first Appeal attempt. It is respectfully requested that the application be allowed to continue to appeal.

1. ***Real Party in Interest.***

The real party in interest in this appeal is E.I. duPont de Nemours and Company, the assignee of the above-referenced patent application.

2. ***Related Appeals and Interferences.***

Application Serial No. 10/414,637 ('637) is being appealed concurrently herewith. The '637 application is a divisional application of the above-referenced application.

3. ***Status of Claims.***

Claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 are the subject of this appeal. Claims 39, 40, 44, 45, 50, and 51 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form. Claims 4-8, 13-16, 20-37, 41, 42, 47, 48, 53, 54, and 65 have been cancelled.

4. ***Status of Amendments.***

No new amendments have been made to the claims.

5. ***Summary of Claimed Subject Matter.***

The pending claims of the present application are drawn to an isolated nucleic acid comprising a nucleotide sequence having at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1, wherein said nucleotide sequence encodes a polypeptide which is pesticidal for at least one pest belonging to the order Coleoptera (see, *e.g.*, page 11, lines 13-18; page 12, lines 4-8; page 18, line 23-30; and page 13, line 10 – page 14, line 9). The claims are also drawn to a transformed plant comprising in its genome at least one stably incorporated nucleotide construct comprising a nucleotide sequence encoding a polypeptide operably linked to a promoter that drives expression of said polypeptide, wherein said polypeptide is pesticidal for at least one pest belonging to the order Coleoptera and wherein said nucleotide sequence has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1 (see, *e.g.*, page and line numbers cited above as well as page 7, line 5 and page 22 lines 3-11). The claims of the present invention are also drawn to a method for impacting an insect pest comprising introducing into a plant or cell thereof at least one nucleotide construct comprising a nucleotide sequence encoding a polypeptide operably linked to a promoter that

drives expression of said polypeptide in plant cells, wherein said polypeptide is pesticidal for at least one pest belonging to the order Coleoptera and wherein said nucleotide sequence has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1, whereby an insect pest feeding on said plant or cell thereof is impacted (see, *e.g.*, page and line numbers cited above as well as page 7, line 6 and page 23, lines 4-12).

Claims 55, 58, and 63 set forth that the percent identity to SEQ ID NO:1 is at least 93%. Claims 56, 59, and 64 require at least 94% and claims 38, 43, and 49 require at least 95% (see, e.g., page 11, lines 13-18 and page 18, lines 23-30).

6. ***Grounds of Rejection to Be Reviewed on Appeal.***

Issue 1—Whether claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 meet the enablement requirement of 35 U.S.C. § 112, first paragraph.

Issue 2—Whether claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 meet the written description requirement of 35 U.S.C. § 112, first paragraph.

7. ***Grouping of Claims.***

Applicants believe that the claims do not stand or fall together. The rejected claims are all "sequence identity claims" which contain limitations that require the nucleotide sequence of the claims to share a specified percent of sequence identity to SEQ ID NO:1. However, the claims differ from each other in the minimum percent sequence identity that is required and may therefore be found to differ in meeting the requirements of patentability. That is, independent claims 1, 9, and 17 and dependent claims 2, 3, 10-12, 18, 19, 46, 52, 57, and 60-62 require that the nucleotide sequence has at least 90% sequence identity to SEQ ID NO:1, while claims 55, 58, and 63 contain limitations requiring the nucleotide sequence to have at least 93% identity to SEQ ID NO:1. Claims 56, 59, and 64 contain limitations requiring the nucleotide sequence to have at least 94% identity to SEQ ID NO:1, and claims 38, 43, and 49 contain limitations requiring the nucleotide sequence to have at least 95% identity to SEQ ID NO:1. While Applicants believe that all these claims are allowable, it is conceivable that among claims with

differing requirements for percent sequence identity, some claims could be found to meet the enablement and written description requirements while others might not. Therefore, the sequence identity claims do not necessarily all stand or fall together. For example, claims requiring at least 90% sequence identity may stand or fall separately from those claims requiring at least 93%, at least 94%, or at least 95% sequence identity.

8. ***Argument.***

(a) Issue 1—Whether claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 meet the enablement requirement of 35 U.S.C. § 112, first paragraph.

The Examiner has rejected the claims under 35 U.S.C. § 112, first paragraph, and indicated that the specification, while being enabling for nucleic acids encoding SEQ ID NOs:2 and 10, expression cassettes comprising the nucleic acids, plants and seeds comprising a construct comprising the nucleic acid, and a method of using it to impact a plant pest, does not reasonably provide enablement for any nucleic acid that has 90% identity to SEQ ID NO:1. For the following reasons, the Examiner's reasoning is not well founded and ignores the guidance provided in the specification and in the art, and the rejection should be reversed.

I. THE CLAIMED INVENTION MEETS THE REQUIREMENTS OF 35 U.S.C. § 112, FIRST PARAGRAPH, FOR ENABLEMENT

To satisfy the requirements of 35 U.S.C. § 112, first paragraph, the specification must teach those skilled in the art to make and use the full scope of the claimed invention without undue experimentation. *Enzo Biochem, Inc. v. Calgene, Inc.*, 188 F.3d 1362, 1371, 52 U.S.P.Q.2d (BNA) 1129, 1135 (Fed. Cir. 1999); *Genentech, Inc. v. Novo Nordisk, A/S*, 108 F.3d 1361, 1365, 42 U.S.P.Q.2d (BNA) 1001, 1004 (Fed. Cir. 1997); *PPG Inds., Inc. v. Guardian Inds. Corp.*, 75 F.3d 1558, 1564, 37 U.S.P.Q.2d (BNA) 1618, 1623 (Fed. Cir. 1996); *In re Wright*, 999 F.2d 1557, 1561-62, 27 U.S.P.Q.2d (BNA) 1510, 1513 (Fed. Cir. 1993); *In re Vaeck*, 947 F.2d 488, 495-96, 20 U.S.P.Q.2d (BNA) 1438, 1444-45 (Fed. Cir. 1991). "That some experimentation may be required is not fatal, the issue is whether the amount of experimentation required is 'undue.'" *In re Vaeck*, 947 F.2d at 495, 20 U.S.P.Q.2d (BNA) at

1444. The enablement section of 35 U.S.C. § 112, first paragraph, "requires that the scope of the claims must bear a reasonable correlation to the scope of enablement provided by the specification to persons of ordinary skill in the art." *In re Fisher*, 427 F.2d 833, 839, 166 U.S.P.Q. (BNA) 18, 24 (C.C.P.A. 1970). In order to determine whether the present claims are enabled, an analysis of the teachings of the specification must be performed as well as an inquiry into the knowledge of persons of ordinary skill in the art. *In re Bowen*, 492 F.2d 859, 861, 181 U.S.P.Q. (BNA) 48, 50 (C.C.P.A. 1974).

A. The Claimed Invention Is Fully Taught in the Specification.

The claims are generally drawn to SEQ ID NO:1 and sequences having identity to SEQ ID NO:1 and in addition require that the nucleotide sequences encode a polypeptide that is pesticidal for at least one pest belonging to the order Coleoptera. The specification teaches those skilled in the art how to make the claimed nucleotide sequences and provides examples of such sequences. The specification provides: nucleotide sequences that fall within the scope of the claims (see, for example, pp. 11, 12, 13, 14, 18, 19, 25, and 65); guidance regarding alterations that allow the amino acid sequence to retain pesticidal activity (see, for example, p. 11, 12, 13, 14, 18 and pp. 19-20); methods for assaying the pesticidal activity of proteins (pp. 8 and 29, Example 4, Example 6, and Example 7); a discussion of Cry-8-like δ-endotoxins (SEQ ID NO:2 is a Cry-8-like δ-endotoxin) (pp. 24-25); guidance for determining percent identity of sequences (pp. 33-38); and, specific mutations that fall within the scope of the claimed invention (pp. 11, 12, 13, 14, 18, 19, and 25; and Examples 4 and 6).

The specification provides multiple truncated variants of SEQ ID NO: 1 and demonstrates that these truncated variants retrain activity. A summary of the sequences presented in the application were presented in Table form for the Examiner's convenience in the Reply Brief filed February 23, 2006 and in the amendment filed August 11, 2006. A copy of the Table is provided as Evidence Appendix A. The Table contains a listing of the sequences set forth in the specification and sequence identity analysis of the sequences with SEQ ID NO:1.

Briefly, SEQ ID NO:1 encodes the amino acid sequence set forth in SEQ ID NO:2. Active variants of SEQ ID NO:1 are set forth in SEQ ID NOS: 3, 15, 19, 11, 23, 31, 33, 29, 9,

43, 21, 39, 41, and, 45. SEQ ID NO:3 comprises a full length active variant of SEQ ID NO:1 and shares 92% sequence identity across the full length of SEQ ID NO:1. The remaining active variants are truncates of SEQ ID NO:1. That is, there are amino acids that are deleted at the 5′ end, the 3′ end, or deletions at both the 5′ and 3′ ends. For each truncated active variant, Table 1 provides both the global percent identity of the variant polynucleotide across the full length of SEQ ID NO:1 and a local percent identity across only the region of the truncate which shares homology to SEQ ID NO:1. Any discrepancy in percent identity numbers in the Table to those previously submitted by Applicants, as noted by the Examiner in the official action of May 18, 2006, merely reflect the rounding up or down of the percent identity numbers.

The alignments summarized in Table 1 were performed using the Emboss Pairwise Alignment Algorithms using the Needle (global) alignment. This software is available from the European Bioinformatics Institute (EMBL-EBI) at http://www.ebi.ac.uk/emboss/align/#. Parameters employed in the protein alignments included a gap open of 10, an extended gap of 0.5 and the Blosum 62 matrix. Parameters employed in the nucleic acid alignments included a gap open of 10, an extended gap of 0.5 and the DNAfull matrix.

The specification provides fourteen (14) active variants of SEQ ID NO:1 which share between 38% and 92% identity across the full length of SEQ ID NO:1. When local alignments are preformed between the truncated active variants and nucleotides 1 to 2097 of SEQ ID NO:1, the percent identity of the active variants to SEQ ID NO:1 ranges between 100% to 68% sequence identity. As multiple active variants have been provided which have a relationship to SEQ ID NO:1 well below the percent identities recited in the instant claims, the claims of the present invention are enabled.

While each of the 14 active variants of SEQ ID NO:1 provided in the specification provides clear support for enablement, a brief discussion of the active variants of SEQ ID NO:9 and 19 are provided to further emphasize the extent to which variants have been enabled. SEQ ID NO:1 comprises 3621 nucleotides. SEQ ID NO:19 comprises 1860 nucleotides of SEQ ID NO:1 and continues to retain activity. The demonstration that such a fragment retains activity clearly illustrates that there is an increased likelihood that an alteration to one or more of the 1,761 additional nucleotides in SEQ ID NO:1 could be altered without disrupting function.

Moreover, the specification provides further evidence that even the 1860 nucleotides which are set forth in SEQ ID NO:19 can be altered and still continue to retain activity. The active variant set forth in SEQ ID NO: 9 is the same truncate as SEQ ID NO:19 except that SEQ ID NO:9 contains maize optimized codons. As shown in the table below, SEQ ID NO:9 shares 38% global sequence identity to SEQ ID NO:1 and shares 68% local sequence identity and continues to retain activity. Accordingly, the data in the specification provides clear guidance to one of skill in the art that active variants having at least 90%, at least 93%, at least 94%, and at least 95% sequence percent identity to SEQ ID NO:1 can be readily made.

B. One of Skill in the Art Can Make and Use the Invention Based on the Teachings of the Specification.

The Federal Circuit has repeatedly stated that enablement is not precluded by the necessity for some experimentation, so long as the experimentation needed to practice the invention is not undue, and that a considerable amount of experimentation is permissible if it is merely routine or if the specification provides a reasonable amount of guidance as to how the experimentation should proceed. *Id. In re Wands*, 858 F.2d 731, 8 USPQ2d 1400 (Fed Cir 1988). In the instant case, the quantity of experimentation required to practice independent claim 1 amounts to two steps: (1) generating a nucleic acid comprising a nucleotide sequence that has at least 90%, at least 93%, at least 94%, or at least 95% sequence identity to SEQ ID NO:1; and (2) assaying the encoded polypeptide for functional activity. Such assays, while known in the art, have further been presented in the specification. See, for example, pages 8 and 29 and in the experimental section in working examples such as Example 4 (pp. 65-66), Example 6 (p. 67), and Example 7 (p. 69). These working examples teach methods for assaying pesticidal activity of proteins and demonstrate results obtained using these assays. One of skill in the art would appreciate that both of these steps are within the skill of those in the art and that this degree of experimentation is not considered undue.

Similarly, the amount of experimentation needed to practice the other sequence identity claims is not undue. For example, independent claim 9 recites a transformed plant comprising a nucleotide construct that has a nucleotide sequence with at least 90% sequence identity to the

nucleotide sequence set forth in SEQ ID NO:1 and that encodes a polypeptide that is pesticidal for at least one pest belonging to the order Coleoptera. Thus, in addition to the steps required to practice independent claim 1, independent claim 9 requires the transformation of a plant. Plant transformation is routine in the art; thus, the amount of experimentation required to practice claim 9 is not undue. Similarly, in addition to the steps required to practice independent claim 1, the method of independent claim 17 requires that a nucleotide construct be created in which the nucleotide sequence is operably linked to a promoter; that the construct be introduced into a plant or cell thereof; and that an insect pest feeding on said plant or cell is impacted. The performance and/or evaluation required by each of these additional steps is within the skill of those in the art and would not be considered undue experimentation by those in the art. Likewise, the remaining sequence identity claims, which are all dependent on or incorporate the limitations of independent claim 1, 9, or 17, contain additional requirements which are equally within the skill of those in the art.

As noted above, SEQ ID NO:3 comprises a full length active variant of SEQ ID NO:1 and shares 92% sequence identity across the full length of SEQ ID NO:1 providing guidance as to residues that can be mutated and still retain activity. Furthermore, the active variant set forth in SEQ ID NO: 9 comprises 1860 nucleotides of SEQ ID NO:1, contains maize optimized codons and continues to retain activity. As noted above, SEQ ID NO:9 shares 38% global sequence identity to SEQ ID NO:1 and shares 68% local sequence identity and continues to retain activity. Therefore, SEQ ID NO:9 demonstrates that the nucleotide sequence can be adjusted using differing codon preferences for individual amino acids to obtain sequences within the scope of the claims.

Applicants note that it was customary in the art at the time of the invention to make and assay a number of sequences for a desired function in order to achieve the best results. For example, common techniques involve what is often referred to as "shuffling," as described for example in U.S. Patent No. 5,837,458, issued November 17, 1998 with inventors Minshull and Stemmer and entitled, "Methods and Compositions for Metabolic and Cellular Engineering." With such techniques, it is common to mutagenize individual sequences or a set of sequences which are then assayed for a desired activity. Such techniques may even make use of a library of

sequences which is recursively mutagenized, screened for function using a functional assay, and re-mutagenized in order to find a sequence exhibiting optimal function. Examples of the use of such techniques include: Minshull and Stemmer (1999) *Current Opinion in Chemical Biology* 3:284-290, entitled "Protein Evolution by Molecular Breeding"; and Christians *et al.* (1999) *Nature Biotechnology* 17: 259-264, entitled "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling." Such experiments are designed and are intended to encompass the generation and testing of a very large number of variant sequences for a desired function. As indicated by these and other publications in the art, this level of experimentation was considered routine in the art and thus would not be considered "undue experimentation" under *In re Wands,* 858 F.2d 731, 8 USPQ2d 1400 (Fed Cir 1988) and *In re Jackson*, 217 USPQ 804, 807 (Bd. Pat. App. & Int. 1982) (holding that a considerable amount of experimentation is permitted to practice the invention and is not undue if it is merely routine in the art or if the specification provides a reasonable amount of guidance and direction to perform such experimentation).

At the time of the invention, it was routine to mutate amino acids in a polypeptide and then test the altered polypeptide for activity. This is proven by the references cited by the Examiner in the previous Office Action. The Lazar reference published in 1988 (Evidence Appendix B) and the Hill reference published in 1998 (Evidence Appendix C), both demonstrate that one of skill in the art well before 2000, the priority date of the present application, could make substitutions in polypeptide sequences and test for activity. Nothing more is required in the present application.

The de Maagd, Tounsi, and Angsuthanasombat references all make substitutions in Cry proteins and then test for activity. This is all that is required to test sequences that fall within the scope of the claims. The de Maagd reference was published in 1999 (Evidence Apendix D) and the Angsuthanasombat was published in 2001 (Evidence Appendix E). Again, these references demonstrate that one of skill in the art could make sequences having at least 90%, at least 93%, at least 94%, or at least 95% sequence identity to a known sequence and test for activity at the time of the invention.

The specification teaches that a comparison of the amino acid sequences of Cry toxins of different specificities reveals five highly conserved sequence blocks. Structurally, the δ-endotoxins comprise three distinct domains, which are, from the N- to C-termini: a cluster of seven alpha- helices implicated in pore formation, three anti-parallel beta sheets implicated in cell binding, and a beta sandwich. *See* page 23 of the specification. The specification further teaches that a truncated protein can be made that retains activity. See pages 23-27 of the specification. The truncated polypeptide would lead one of skill in the art to conclude that the deleted region would tolerate modifications. Furthermore, one of skill in the art would appreciate that changes would be more likely to be tolerated outside of conserved domains. Thus, there is teaching in the specification that would guide one of skill in the art in making modifications. The specification further teaches preparing modified sequences and testing such sequences for activity. *See*, for example, pages 23, 24, and 29 as well as Examples 1, 6, and 7 of the Specification, and Table 1, above.

As assays for determining whether the modified sequences would retain activity were disclosed, one of skill in the art as of the filing date of the present application would have been able to make such modifications and test them for pesticidal activity. Nothing more is required to fully enable the claims. Accordingly, one of skill in the art would be able to determine the functionality of polypeptides encompassed by the claimed invention without resorting to undue experimentation and therefore the enablement requirement is satisfied.

In fact, Applicants note that Dr. André Abad submitted a Rule 132 declaration with the Amendment of 1/21/05 stating, among other things, that the procedures described in Examples 4, 6, and 7 of the specification are considered "routine." The Rule 132 declaration was filed with Applicant's amendment of January 21, 2005 as Appendix C and is included with the Brief as Evidence Appendix F. The Examiner has dismissed Dr. Abad's declaration as "only provid[ing] opinions, and ones the Declarant is not even sure of" (Office Action of 5/6/05, page 8, second and third paragraphs). Applicants believe that this summary by the Examiner is inaccurate and improper. The above-cited references, (Lazar, Hill, de Maagd, Tounsi, and Angsuthanasombat) which disclose mutating amino acid sequences and assaying for activity, all support Dr. Abad's declaration and prove that the present claims are fully meet the enablement standard.

C. The Examiner's Reasoning Is Not Well-Founded.

In contrast to the conclusions stated in the Office Action (*e.g.*, page 4, first full paragraph), guidance is provided as to what sequence alterations may be made and still provide a polypeptide species encompassed by the claim. Applicants have provided the exemplary nucleotide sequence of SEQ ID NO: 1 and the exemplary amino acid sequence of SEQ ID NO: 2. The claimed sequences of the invention vary from this sequence by structural parameters (*i.e.*, percent sequence identity to SEQ ID NO: 1). Guidance for determining percent identity of sequences is provided in the specification on pages 33 through 38.

Moreover, independent claims 1, 9, and 17, in addition to requiring a structural component (at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 1), specify that the nucleotide sequence encodes a polypeptide which is pesticidal for at least one pest belonging to the order Coleoptera and therefore these claims (and claims dependent thereon) encompass functional variants. Guidance regarding alterations that allow the sequence to retain the specified pesticidal activity is also provided. See, for example, page 18, line 17 through page 19, line 20 that discuss variants and their pesticidal activity. In addition, methods for assaying pesticidal activity of proteins are routine in the art and are also described in the specification, for example on page 8, lines 24-30 and on page 28, lines 24-30 and in the experimental section in working examples such as Example 4, (pp. 65-66), Example 6 (pp. 67-69), and Example 7 (pp. 69-73).

The Examiner is referred to Table 1 (Evidence Appendix A). As explained above, the Table provides numerous examples of sequences, all of which are provided in the specification, some having less than the required sequence identity and having the required activity. Thus, the claims are enabled.

On page 6 of the Office Action, lines 3-10, the Examiner reasons that to enable the claims, one must make and test all possible combinations of nucleic acids falling within the scope of the claim.

> Making all possible single amino acid substitutions, in an 3621 nucleotide long nucleic acid like that of SEQ ID NO:1 would require making and analyzing

> 19^{3621} nucleic acids; these nucleic acids would have about 99.99% identity to SEQ ID NO:1. Because nucleic acids that have 90% identity to SEQ ID NO:1 would have up to 362 nucleotide substitutions, many more than 19^{3621} nucleic acids would need to be made and analyzed.

The Examiner's analysis is improper. As held by the court in *In re Borkowski*, 422 F.2d 904, 909, 164 U.S.P.Q. (BNA) 642, 645 (C.C.P.A. 1970), it is inappropriate "to study appellants' disclosure, to formulate a conclusion as to what he (the Examiner) regards as the broadest invention supported by the disclosure, and then to determine whether appellant's claims are broader than the Examiner's conception of what 'the invention' is." In the present case, the methods and examples disclosed in the specification readily teach one of skill in the art to make and test sequences having at least 90% identity to SEQ ID NO:1.

The specification provides guidance to one of skill in the art for making modifications, describes the domains of the Cry protein, and provides insights as to where modifications may be tolerated. See, for example, pages 23 and 24. The specification further teaches preparing modified sequences and testing such sequences for activity. See, for example, pages 23, 24, and 29, as well as, Examples 1, 6, and 7. Modified versions and truncated versions of the polypeptide are disclosed retaining pesticidal activity. See, for example, Table 1 provided herein.

The Examiner also ignores the information available in the art as of the filing date regarding δ-endotoxins. δ-endotoxins are extremely well-characterized and related to various degrees by similarities in their amino acid sequences and tertiary structures. The specification contains a reference to Li *et al.* (Evidence Appendix G) as well as a discussion on designing mutant sequences. *See*, for example, page 25 of the specification:

> The inventors of the present invention used the solved structure of the *Cry3A* gene (Li *et al.* (1991) *Nature* 353:815-821) to produce a homology model of the Cry8 δ-endotoxin disclosed and claimed herein as SEQ ID NO:2 to gain insight into the relationship between structure and function of the endotoxin, and to design the recombinantly engineered proteins disclosed and claimed herein. A combined consideration of the published structural analyses of *B. thuringiensis* endotoxins and the reported function associated with particular structures, motifs, and the like indicates that specific regions of the endotoxin are correlated with particular functions

> and discrete steps of the mode of action of the protein. For example, δ-endotoxins isolated from *B. thuringiensis* are generally described as comprising three domains, a seven-helix bundle that is involved in pore formation, a three-sheet domain that has been implicated in receptor binding, and a beta-sandwich motif (Li *et al.* (1991) *Nature,* 305:815-821).
>
> The inventors reasoned that the toxicity of Cry8-like proteins, and specifically the toxicity of the Cry8 protein, could be improved by targeting the region located between alpha helices 3 and 4 of domain 1 of the endotoxin protein. This theory was premised both on the knowledge that alpha helices 4 and 5 of domain 1 of Cry3A δ-endotoxins had been reported to insert into the lipid bilayer of cells lining the midgut of susceptible insects (Gazit *et al.*, (1998) *PNAS USA* 95:12289-12294); the inventors' knowledge of the location of trypsin and chymotrypsin cleavage sites within the amino acid sequence of the wild-type protein; and the observation reported herein that the protein encoded by 1218-1 (i.e., SEQ ID NO:2) was more active against certain Coleopterans following *in vitro* activation by trypsin or chymotrypsin treatment. Accordingly, the inventors engineered a mutant Cry8-like protein that would comprise at least one additional trypsin cleavage site in the region located between helices 3 and 4 of domain 1.

Specification, p. 25, lines 1-24.

Yet, even in view of the description provided in the specification, the many examples taught, and the knowledge available in the art, the Examiner ignores all the teachings. The specification provides truncated polypeptides containing additional modifications that retain activity. Additionally, SEQ ID NO:3 is provided which is a natural variant of SEQ ID NO:1.

The Examiner discounts all the disclosure and teachings and provides only flawed reasoning for her conclusion. The references cited by the Examiner support Applicants' position that the claims are enabled.

Under the facts of the present application, one skilled in the art would understand whether a particular protein has at least 90%, 93%, 94%, or 95% sequence identity with SEQ ID NO:1 as set forth in the claims. In addition, functional assays are disclosed in the specification and provide sufficient guidance for one skilled in the art to determine whether a particular polynucleotide is within the scope of the claims. Thus, the claims are fully enabled.

At pages 7 and 8 of the Office Action, the Examiner lists several objections.

First, the Examiner notes that Table 1 has numerous errors. As noted above, Applicants have deleted the "Description" column of the table to avoid any confusion. The more important information in the Table is the % identity that the sequences share with the claimed sequences. The Examiner has further noted that in a previous response, SEQ ID NO:19 was indicated as having different percent global identity to SEQ ID NO:1. In the previous response, the percent identity was listed as 51% whereas 52% is indicated in Table 1. The difference is simply a reflection of rounding off of the numbers. However, regardless of whether it is 51% or 52%, the variant clearly supports enablement for claims having at least 90% identity to SEQ ID NO:1.

Second, the Examiner notes that some of the sequences are truncated and that one of skill in the art would only look at it as a truncated protein. The Table provides % identity of the truncated nucleotide sequences and proteins across the full length of SEQ ID NO:1 and SEQ ID NO:2 as well as across the corresponding nucleotides of SEQ ID NO:1 and amino acids of SEQ ID NO:2. Thus, the truncated sequences establish that these is a region that can be deleted and yet the protein retains activity. One of skill in the art would recognize that this region would likely tolerate modifications in the deleted region. Further, as noted, even within the corresponding region, changes are tolerated.

It is further noted that the Examiner on page 4 of the Official Action states: "[t]he specification also fails to provide guidance for which amino acids can be deleted and which regions of the proteins can tolerate insertions and still produce a function enzyme." The Examiner is clearly disregarding the examples provided in the application. See Evidence Appendix A.

Third, the Examiner notes that the specification does not consider deletions the same as substitutions. First, the claim does not require a deletion or a substitution. The claims are drawn to sequences having at least 90%, 93%, 94% or 95% sequence identity with SEQ ID NO:1. Numerous examples are set forth in the specification as discussed above. Secondly, the Examiner does not consider SEQ ID NO:3 which is a full length variant of SEQ ID NO:1 and has 92% sequence identity to SEQ ID NO:1.

Fourth, the Examiner would require Applicants to make and test every possible substitution. This is improper and counter to Federal Circuit and PTO Board authority for the

Examiner to require Applicants to demonstrate every substitution that could be made in the sequence. While the claims encompass more, many changes can be made in the nucleotide sequence that would not change the encoded amino acid sequence. The Examiner is directed to SEQ ID NO:9 which shares only 68% local sequence identity and continues to retain activity.

Fifth, the Examiner indicates that SEQ ID NO:3 cannot provide support because "the vast majority of substitutions are localized to such a small area, the protein it encodes, SEQ ID NO:4 has 89% identity to SEQ ID NO:2. Thus, SEQ ID NO:3 cannot provide guidance for making nucleic acids with 90% identity to SEQ ID NO:1 and encoding a protein with 362 amino acid substitutions relative to SEQ ID NO:2" Page 8, lines 14-17. First, SEQ ID NO:3 provides support for substitutions that can be made to the nucleic acid sequence and the resulting protein retain activity. Secondly, again the Examiner is misguided in requiring that Applicants demonstrate all the possible substitutions to provide enablement.

D. That Some Experimentation May Be Necessary Does Not Indicate That the Claims Are Not Enabled.

It is recognized that in unpredictable art areas, the court has refused to find broad generic claims enabled where the corresponding specifications only demonstrate the enablement of one or very few embodiments and do not demonstrate with reasonable specificity how to make and use other potential embodiments across the full scope of the claim. *See, e.g., In re Goodman*, 11 F.3d 1046, 1050-52, 29 U.S.P.Q.2d (BNA) 2010, 2013-15 (Fed. Cir. 1993); *Amgen, Inc. v. Chugai Pharm. Co.*, 927 F.2d 1200, 1212-14, 18 U.S.P.Q.2d (BNA) 1016, 1026-28 (Fed. Cir. 1991); *In re Vaeck*, 947 F.2d at 496, 20 U.S.P.Q.2d (BNA) at 1445. The court has explained that enablement is lacking in those cases because the undescribed embodiment cannot be made based on the disclosure in the specification, without undue experimentation. However, the court has made clear that the question of undue experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is that the amount of experimentation "must not be unduly extensive." *Atlas Powder Co. v. E.I. DuPont de Nemours & Co.*, 750 F.2d 1569, 1576, 224 U.S.P.Q. (BNA) 409, 413 (Fed. Cir. 1984). The Patent and Trademark Office Board of Appeal has indicated: "the test is not merely quantitative,

since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed." *Ex Parte Jackson*, 217 U.S.P.Q. (BNA) 804, 807 (1982).

In the present case, all the Examiner has established is that some experimentation would be required to make and use other embodiments of the claimed invention. What the Examiner has not done is perform the fact-finding needed in order to reach a proper conclusion of undue experimentation. The Examiner has not relied upon any evidence in support of this rejection which would establish that making and testing other sequences beyond those described in the present specification amounts to undue experimentation. In fact, the Examiner has ignored the guidance in the specification, the presence of working examples, and the teachings of the prior art. The references cited by the Examiner support the position that the procedures described in Examples 4, 6, and 7 for making and testing modified sequences are routine in the art. The Examiner makes the rejection based upon unsupported opinions.

E. The Examiner Mischaracterizes the Cited References.

The Examiner argues that making conservative amino acid substitutions does not produce predictable results and cites Lazar *et al.*, *Molecular & Cellular Biology* 8:1247-1252 (1988) in support of her position. The Examiner indicates that the conservative substitution of glutamic acid for aspartic acid at position 47 reduced biological function of transforming growth factor alpha while nonconservative substitutions with alanine or asparagine had no effect. The Examiner fails to consider the entire teachings of the reference.

First, the Lazar reference is drawn to studying transforming growth factor α (TGF-α). TGF-α is a mammalian polypeptide of 50 amino acids. The polypeptide is in no way related to the Cry proteins of the present invention. The reference relating to TGF-α does not bear any relevance to the claimed Cry proteins.

Secondly, with respect to the modifications described by Lazar, two amino acids of TGF-α which were known to be conserved among the family of EGF-like polypeptides were modified.

It would come as little surprise to one skilled in the art that the modification of such a conserved amino acid should lead to the loss of function described by the authors. One of the changes at position 47 described by the authors indicates that [Asn-47]- TGF-α retains biological activity. The authors note that interestingly, two of the EGF-like viral proteins, myxomal growth factor and Shope fibroma growth factor, have Asn instead of Asp in position 47. Thus, the reference supports Applicants' position that protein domains are important and that by aligning sequences, one of skill in the art can determine what sites would likely tolerate changes.

On page 11 of the Office Action, the Examiner misinterprets the teachings of the Lazar reference. The purpose of the Lazar study was NOT to test "the commonly accepted notion that making conservative substitutions would not affect activity and that nonconservative substitutions would affect activity, even inactivate the protein." As explicitly set forth in the abstract, page 1247, the purpose of the Lazar work was "[t]o study the relationship between the primary structure of transforming growth factor α (TGT- α) and some of its functional properties." To do so, conservative and nonconservative amino acid substitutions were made at two amino acid positions that are highly conserved. The results are not relevant to substitutions outside of conserved regions.

The Examiner additionally cites Hill *et al.*, *Biochemical & Biophysical Research Communications* 244:573-577 (1998) as supporting the position that substitution of a residue with a conservative amino acid can drastically reduce enzyme activity. The Examiner cites Hill as teaching that when three histidines that are maintained in ADP-glucose pyrophosphorylase across several species are substituted with the 'nonconservative' amino acid glutamine, there is little effect on enzyme activity, while the substitution of one of those histidines with the 'conservative' amino acid arginine drastically reduced enzyme activity.

First, the Hill reference is drawn to studying ADP-glucose pyrophosphorylase. The polypeptide is in no way related to the Cry proteins of the present invention. The reference relating to ADP-glucose pyrophosphorylase does not bear any relevance to the claimed Cry proteins.

Secondly, with respect to the modifications described by Hill, the modified residues were conserved among bacterial and plant ADP-glucose pyrophosphorylases. As set forth in the first

line of the abstract, "[t]wo **absolutely conserved** histidines and a third **highly conserved** histidine are noted in 11 bacterial and plant ADP-glucose pyrophosphorylases." (emphasis added) These **absolutely** and **highly conserved** histidines were mutagenized and characterized in the paper. It would come as little surprise to one skilled in the art that the modification of one of these conserved amino acids should lead to the loss of function described by the authors.

The Examiner states on page 9, lines 15-16, that "hill *et al.* [*sic*] teach that conserved blocks cannot be relied on in making amino acid substitutions." The Examiner is completely mischaracterizing the reference. Even Hill states, "[c]omparisons of these sequences highlight those strictly conserved residues whose functions are essential." Hill *et al.* Page 573, column 2. The paper further notes that while the three conserved histidines are not essential for catalytic activity, when a substitution is made, activity is slightly lower than wild type. Page 576.

The paper does not speak to changes outside of conserved regions as taught in the present application.

The Examiner indicates that making amino acid substitutions in Cry proteins is unpredictable and cites three references drawn specifically to Bt toxins, de Maagd *et al.* (1999) *Appl. Environ. Microbiol.* 65:4369-4374; Tounsi *et al.* (2003) *J. Appl. Microbiol.* 95:23-28; and Angsuthanasombat *et al.* (2001) *J. Biochem. Mol. Biol.* 34:402-407. As discussed below, the Examiner misinterprets the references. When taken as a whole, the references support Applicants' arguments that one of skill in the art could make and use the invention.

The de Maagd reference is drawn to the identification of Cry1C domain III amino acid residues involved in insect specificity. As the article explains, progress has been made both in determining the three-dimensional structure of the toxin molecule and in identifying the primary sequences involved in specificity and receptor binding, allowing the study of structure-function relationships (page 4369, column 1). Domain III was known to be involved in binding of toxins to putative receptors of brush border membranes of insects. The study was undertaken to identify amino acid residues in domain III involved in specificity for beet armyworm. As noted in the abstract, the results "identify groups of amino acids as well as some individual residues in Cry1C domain III, which are strongly involved in *S. exigua*-specific activity as well as sometimes involved in *M. sexta*-specific activity." The work aligned domain III from Cry1C,

Cry1E, and Bs21 and showed that only blocks B through E of Cry1C or parts thereof, are essential for a high level of activity against *S. exigua*. The work described in the publication was an effort to determine which parts of domain III of Cry1C are involved in *S. exigua*-specific activity. Individual amino acids that are involved in the specificity were identified by mutagenesis. The paper proves that with the knowledge in the art on Cry endotoxins, specific amino acids involved in binding specificity can be identified. Likewise, with this same knowledge, mutations can be made in the amino acid sequence and the resulting sequences tested for those that retain activity. This work supports Applicants' arguments that changes can be made and function preserved.

The Tounsi reference reports on the study of the expression of a new Cry11a-type gene. By comparing the sequence of the new gene with known Cry11a genes, amino acid differences were detected (Figure 2, page 26). The paper notes that these substitutions may be important for studies on toxicity. Again, the paper supports the principle that using alignments and domain information substitutions can be made in Cry proteins and activity retained.

The Angsuthanasombat paper reports on the directed mutagenesis of a Cry11a toxin. The paper notes that α4 and α5 of the 130kDa Cry 4B toxin are essential determinants of toxicity. Using amino acid sequence alignment with Cry1Aa and Cry3A and the homology model of Cry4B, the predicted α4 and α5 were located. To investigate the possible role for toxicity of charged and polar amino acids in α4 of Cry11A, eight Cry11A mutants were generated. As set forth on pg 405, only the R136A mutation resulted in total loss of larvicidal activity.

It is not surprising that a mutation in a critical region known to be essential for toxicity would have an effect on toxicity. The seven mutations did not abolish activity, but did affect the level of activity.

F. The Specification Meets the Federal Circuit Standard for Enablement and the Rejection Should Be Withdrawn.

The Federal Circuit has repeatedly stated that enablement is not precluded by the necessity for some experimentation, so long as the experimentation needed to practice the invention is not undue. *In re Wands* 8 USPQ2d 1400 (Fed Cir 1988). Furthermore, a

considerable amount of experimentation is permissible, if it is merely routine, or if the specification provides a reasonable amount of guidance in which the experimentation should proceed. *Id.*

Applicants stress that when evaluating the quantity of experimentation required, the court looks to the amount of experimentation required to practice a single embodiment of the invention, rather than the amount required to practice every embodiment of the invention, as the Office Action implies. For example, in *Wands*, the claims at issue were drawn to immunoassay methods using any monoclonal antibody having a binding affinity for HbsAg of at least 10^{-9} M. The PTO had taken the position that the claim was not enabled because undue experimentation would be needed to make the monoclonal antibodies required for the assay. The Federal Circuit reversed and held that the claims were enabled, as the amount of experimentation required to isolate monoclonal antibodies and screen for those having the correct affinity was not undue. *See Id.* Clearly, the Federal Circuit did not contemplate that every antibody useful in the methods of the claim must be identified. Rather, the court considered the amount of experimentation required to identify one or a few monoclonal antibodies having the required affinity. *See also, Johns Hopkins University v. Cellpro*, 931 F. Supp. 303, 324 (D. Del. 1996), *aff'd in part, vacated in part, and remanded*, 47 USPQ2d 1705 (Fed. Cir. 1998) (stating that "[t]he specification need only enable one mode of making the claimed invention.").

In the instant case, the quantity of experimentation required to practice independent claim 1 amounts to two steps: generating a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 1 and assaying the encoded polypeptide for pesticidal activity against at least one pest belonging to the order Coleoptera. Such assays, while routine in the art, have further been presented in the specification. Similarly, the amount of experimentation needed to practice the other claims is not undue. For example, claim 9 recites a transformed plant comprising a nucleotide construct comprising a nucleotide sequence that has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 1 and that encodes a polypeptide that is pesticidal for at least one pest belonging to the order Coleoptera. Thus, in addition to the steps required to practice claim 1, claim 9 requires the transformation of a plant. Plant transformation

is routine in the art and is also readily achieved by those of skill in the art. Thus, a rational scheme for practicing the claimed invention is provided.

Based on the guidance regarding the exemplary nucleotide and polypeptide sequences of the invention and the methods for determining whether a particular polypeptide has pesticidal activity against at least one insect of the order Coleoptera, the skilled artisan could choose among possible sequence modifications to produce polypeptides within the parameters set forth in the claims and then test these sequence variants to determine if they retained pesticidal activity. Consequently, contrary to the conclusions stated in the Office Action, the quantity of experimentation necessary and the amount of guidance presented in the specification is sufficient to enable the claims. In view of this discussion, Applicants respectfully request that the rejection of claims under 35 U.S.C. §112, first paragraph, be withdrawn.

Moreover, Applicants note that recent Board of Appeals decisions support Applicants' arguments that the present claims meet the enablement requirement. See *Ex parte Sun*, No. 2003-1993 (Bd. Pat. App. Int., Jan. 20, 2004) and *Ex parte Vogelstein*, No. 2002-0779 (Bd. Pat. App. Int., Dec. 30, 2002). While these decisions are specifically indicated to not be precedential, Applicants note that similar enablement rejections of similar claims were reversed by the Board in view of similar support.

In establishing nonenablement, the burden rests initially with the Examiner to substantiate the unpredictability of the art and that, given the unpredictability, the specification does not provide sufficient information to guide those of skill to make and use the claimed invention across the full scope of the claims. In the present case, a clear goal is disclosed. Furthermore, guidance is provided for making the claimed sequences, assays are provided to determine whether modified sequences would encode proteins that retain activity, examples are provided showing that modifications to the nucleotide sequence can be made and the encoded proteins retain pesticidal activity, and art is cited that provides information on the Cry proteins of the invention. Thus, whatever unpredictability surrounds the construction of other sequences, the need for undue experimentation is mitigated by the examples of how to make and use such claimed sequences.

The Examiner repeatedly argues that the specification does not teach which amino acid substitutions can be made to retain pesticidal activity and that while the claims require that the nucleic acid encode pesticidal proteins the specification does not teach how to make such nucleic acids. These statements do not take into account the numerous truncated and mutant sequences set forth in the specification. As the Federal Circuit has noted, "That some experimentation may be required is not fatal, the issue is whether the amount of experimentation required is 'undue.'" *In re Vaeck*, 947 F.2d at 495, 20 U.S.P.Q.2d (BNA) at 1444.

For all these reasons, the rejection of the claims under 35 U.S.C. §112, first paragraph as lacking enablement should be withdrawn.

G. Claims 38, 43, 49, 55, 56, 58, 59, 63, and 64 Should Be Considered Separately.

Claims 55, 58 and 63 require the nucleotide sequence to have 93% sequence identity to SEQ ID NO:1. Claims 56, 59, and 64 require the nucleotide sequence to have 94% sequence identity to SEQ ID NO:1. Claims 38, 43, and 49 require the nucleotide sequence to have 95% sequence identity to SEQ ID NO:1.

Applicants submit that the disclosure in the application, the examples provided in the application, the knowledge in the art on endotoxin genes, and the methods for assaying for activity are sufficient to enable all of the claims. However, should the Board disagree, claims requiring 93%, 94%, and 95% sequence identity are narrower in scope and thus should be considered in view of the arguments presented above.

(b) Issue 2—Whether claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 meet the written description requirement of 35 U.S.C. § 112, first paragraph.

Claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 were rejected under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement. This rejection is respectfully traversed. For the reasons set forth below, the rejection of the claims under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement should be withdrawn.

II. THE CLAIMED INVENTION MEETS THE REQUIREMENTS OF 35 U.S.C. § 112, FIRST PARAGRAPH, FOR WRITTEN DESCRIPTION

An adequate written description for genetic material requires a precise definition, "such as by structure, formula, chemical name, or physical properties." *See Enzo Biochem Inc. v. Gen-Probe, Inc.*, 323 F.3d 956, 962-963, 970 (Fed. Cir. 2002). The goal of the written description requirement is to clearly convey that an applicant has invented the subject matter which is claimed. *See*, for example, *In re Barker*, 559 F.2d 588, 592, (CCPA 1977). To satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonable conclude that the inventor had possession of the claimed invention. *See*, *Moba, B.V. v. Diamond Automation, Inc.* 325 F.3d 1306, 1319, (Fed. Cir. 2003); and *Vas-Cath, Inc. v. Mahurkar*, 935 F2d 1555, 1563 (Fed. Cir. 1991). The written description inquiry focuses on whether the specification reasonably conveys to one skilled in the art whether the applicant invented the claimed subject matter. Thus, the relevant inquiries are: What is the applicant's claimed invention? What is now claimed?

A. The Claimed Sequences Are Adequately Described in the Specification.

The claimed invention is directed to nucleotide sequences having specific structural and biological properties. The specification provides both the DNA and amino acid sequences of a representative embodiment of the claimed sequences. Indeed, the Examiner has acknowledged that these claims drawn to specific sequences would be allowable if rewritten as independent claims. The specification also discloses modified sequences that fall within the scope of the claims. *See*, for example, Evidence Appendix A. Accordingly, the application provides the structural features that characterize nucleic acid sequences having 90%, 93%, 94%, or 95% identity to SEQ ID NO:1 and still retain pesticidal activity. The sequences that fall within the scope of the claims can readily be identified by the methods set forth in the specification.

Table 1 provided as Evidence Appendix A is a guide of the sequences disclosed in the specification. As one can see from the Table, one of skill in the art would readily understand that

Applicants were in possession of the claimed invention. The Table provides sequences that are modified yet still retain activity. Accordingly, the claims are fully described in the specification.

Applicants note that the description of a representative number of species does not require the description to be of such specificity that it would provide individual support for each species that the genus embraces. 66 Fed. Reg. 1099, 1106 (2001). Satisfactory disclosure of a "representative number" depends on whether one of skill in the art would recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus in view of the species disclosed. 66 Fed. Reg. 1099, 1106 (2001). Applicants submit that the knowledge and level of skill in the art would allow a person of ordinary skill to envision the claimed invention, *i.e.*, a nucleotide sequence having at least 90% sequence identity to the sequence set forth in SEQ ID NO: 1.

The description of a claimed genus can be by structure, formula, chemical name, or physical properties. *See, Ex parte Maizel*, 27 USPQ2d 1662, 1669 (B.P.A.I. 1992), citing *Amgen v. Chugai*, 927 F.2d 1200, 1206 (Fed. Cir. 1991). A genus of DNAs may therefore be described by means of a recitation of a representative number of DNAs defined by nucleotide sequence and falling within the scope of the genus, *or* by means of a recitation of structural features common to the genus, which features constitute a substantial portion of the genus. *See, Regents of the University of California v. Eli Lilly & Co.*, 119 F.3d 1559, 1569 (Fed. Cir. 1997) (referred to herein as "*Lilly*"); *see also* Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, First Paragraph, "Written Description" Requirement, 66 Fed. Reg. 1099, 1106 (2001) (referred to herein as the "Guidelines"). The Office Action cites *Lilly* but ignores this aspect of the case as well as the Guidelines. Indeed, the passages cited in the Office Action refer to the description of a sequence by function alone without any structural limitation, which is not the case here. All of the pending claims recite a functional limitation and also require a predictable structure of at least 90% sequence identity to SEQ ID NO: 1. Under both Lilly and the Guidelines, these requirements for function in combination with the recitation of a predictable structure should be sufficient to satisfy the written description requirement.

Applicants note that the Federal Circuit has explicitly stated that

> *Eli Lilly* did not hold that all functional descriptions of genetic material necessarily fail as a matter of law to meet the written description requirement; rather, the requirement may be satisfied if in the knowledge of the art the disclosed function is sufficiently correlated to a particular, known structure.

Amgen, Inc. v. Hoechst Marion Roussel, Inc., 314 F.3d 1313, 1332 (Fed. Cir. 2003). *See also, Moba, B.V. v. Diamond Automation, Inc.*, 325 F.3d 1306, 1320 (noting that "[i]n more recent cases, however, this court has distinguished *Lilly*" and further noting that in *Enzo Biochem, Inc. v. Gen-Probe, Inc.*, 323 F.3d 956 (Fed. Cir. 2002), "neither the specification nor the deposited biological material recited the precise 'structure, formula, chemical name, or physical properties' required by *Lilly*.")

Example 14 of the "Synopsis of Application of Written Description Guidelines" is directed to a generic claim: a protein having at least 95% sequence identity to the sequence of SEQ ID NO: 3, wherein the sequence catalyzes the reaction A→B. The synopsis materials conclude that the generic claim of Example 14 is sufficiently described under §112, first paragraph, because: 1) "the single sequence disclosed in SEQ ID NO: 3 is representative of the genus"; and 2) the claim recites a limitation requiring the compound to catalyze the reaction from A→B. The synopsis materials conclude that one of skill in art would recognize that the Applicants were in possession of the necessary common attributes possessed by the members of the genus.

Following the analysis of Example 14, Applicants submit that the present claims satisfy the written description requirements of § 112, first paragraph. Specifically, the claims of the present invention encompass sequences having at least 90% sequence identity to SEQ ID NO: 1, wherein the encoded polypeptide is pesticidal for at least one pest belonging to the order Coleoptera. As in Example 14, the specification discloses the nucleic acid sequence of SEQ ID NO: 1 and claims recite a limitation requiring the compound to have a specific function (*i.e.*, pesticidal activity). Consequently, contrary to the conclusion stated in the Office Action, the sequences encompassed by the claims are defined by relevant identifying physical and chemical properties. In fact, the common attributes or features of the elements possessed by the members of the genus is that they encode polypeptides having pesticidal activity against at least one pest

of the order Coleoptera and share at least 90% sequence identity at the nucleotide level to the disclosed nucleotide sequence of SEQ ID NO:1. The necessary common features of the claimed genus are clear.

In summary, the description of a representative number of species *does not* require the description to be of such specificity that it would provide individual support for each species that the genus embraces. Applicants submit that the relevant identifying physical and chemical properties of the disclosed genus would be clearly recognized by one of skill in the art and consequently, the Applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus. Accordingly, the rejection of claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 under 35 U.S.C. §112, first paragraph, for lack of written description should be withdrawn.

B. The Examiner Does Not Consider the Written Description Support in the Specification.

As set forth above, there is adequate written description support in the specification for the claims. The Examiner argues that "[t]he only species reduced to practice in the specification is SEQ ID NO:1, which encodes SEQ ID NO:2." The Examiner is overlooking SEQ ID NOs: 3, 15, 19, 11, 23, 31, 33, 29, 9, 43, 21, 39, 41, and 45. See Table 1, included above. The Table includes the overall % identity and the local % identity of the sequences with SEQ ID NO:1. As set forth in the Table, SEQ ID NOs: 1 and 3 represent sequences having 92% sequence identity. Additionally, the specification provides multiple truncated variants and demonstrates that these truncated variants retrain activity. These sequences demonstrate that sequences having at least 90% sequence identity to SEQ ID NO:1 can be made and retain activity. As Table 1 demonstrates, variant sequences not only can be made that retain activity but are disclosed within the specification. A summary of the sequences presented in the application are provided in Table 1. Briefly, SEQ ID NO:1 encodes the amino acid sequence set forth in SEQ ID NO:2. Active variants of SEQ ID NO:1 are set forth in SEQ ID NOS: 3, 15, 19, 11, 23, 31, 33, 29, 9, 43, 21, 39, 41, and, 45. SEQ ID NO:3 comprises a full length active variant of SEQ ID NO:1 and shares 93% sequence identity across the full length of SEQ ID NO:1. The

remaining active variants are truncates of SEQ ID NO:1. For each truncated active variant, Table 1 provides both the global percent identity of the variant polynucleotide across the full length of SEQ ID NO:1 and a local percent identity across only the region of the truncate which shares homology to SEQ ID NO:1.

The specification provides fourteen (14) active variants which share between 38% and 92% identity across the full length of SEQ ID NO:1. When local alignments are performed between the truncated active variants and nucleotides 1 to 2007 of SEQ ID NO:1, the percent identity of the active variants to SEQ ID NO:1 ranges between 68% to 100% sequence identity. As multiple active variants have been provided which have a relationship to SEQ ID NO:1 well below the percent identities recited in the instant claims, the written description requirement is met.

The Examiner's statement that "Applicants have [sic] not, in fact, described nucleic acids with 90%, 93%, 94% or 95% identity to SEQ ID NO:1 and that encode a protein pesticidal for at least one pest belonging to the order Coleoptera within the full scope of the claims, and the specification fails to provide an adequate written description of the claimed invention." is incorrect. Page 14, lines 16-19, of the Office Action. Table 1 and the 14 sequences contained therein, all of which are set forth in the specification, indicate otherwise. Accordingly, the claims are fully described in the specification.

C. The Facts of the Present Case Are Distinguishable from *Lilly* and *Fiers*.

The Examiner quotes *Eli Lilly* at page 1406 stating "A [*sic*] description of a genus of cDNAs may be achieved by means of a recitation of a representative number of cDNAs, defined by nucleotide sequence. . ." First, as noted above, the specification provides more information than merely reciting SEQ ID NO:1. The specification provides: nucleotide sequences that fall within the scope of the claims (see, for example, pp. 11, 12, 13, 14, 18, 19, 25, and 65 and Table 1), guidance regarding alterations that allow the amino acid sequence to retain pesticidal activity (see, for example, p. 18 and pp.19-20); methods for assaying the pesticidal activity of proteins (pp. 8 and 29, Example 4, Example 6, and Example 7); a discussion of Cry-8-like δ-endotoxins (SEQ ID NO:2 is a Cry-8-like δ-endotoxin) (pp. 24-25); guidance for determining percent

identity of sequences (pp. 33-38); and, specific mutations that fall within the scope of the claimed invention (Example 4 and Example 6).

Secondly, the Examiner's appeal to *Eli Lilly* is misplaced. As noted by the Federal Circuit in *Invitrogen Corp. v. Clontech Laboratories, Inc.* 77 U.S.P.Q.2d (BNA) 1161, 1175 (Fed. Cir. 2005), "[i]n those cases . . . , *University of California v. Eli Lilly & Co.*, 119 F.3d 1559, 43 U.S.P.Q.2d (BNA) 1398 (Fed. Cir. 1997) and *Fiers v. Revel*, 984 F.2d 1164, 1171, 25 U.S.P.Q.2d (BNA) 1601, 1606 (Fed. Cir. 1993), the patent specifications at issue did not identify the sequence (structure) of any embodiment of DNA claimed therein (citations omitted). In contrast, the shared written description for the patents-in-issue recites both the DNA and amino acid sequences of a representative embodiment of the claimed RT enzyme. The specification also discloses test data that the enzyme produced by the listed sequence has the claimed features--DNA polymerase activity without RNase H activity. Under both the *Eli Lilly* and *Fiers* analysis, the specification at bar is sufficient." *Id.* at 1073.

In the present application, a representative nucleic acid and amino acid sequence is provided. Additionally, modified sequences are disclosed which are representative of the claimed sequences. Methods for assaying for activity are provided. Accordingly, the application meets the requirement for written description for the claimed sequences.

In view of the above arguments and amendments, all grounds for rejection under 35 U.S.C. §112, first paragraph, have been overcome. Accordingly, it is respectfully submitted that the rejections under 35 U.S.C. §112, first paragraph, should be withdrawn.

D. Claims 38, 43, 49, 55, 56, 58, 59, 63, and 64 Should Be Considered Separately.

Claims 55, 58 and 63 require the nucleotide sequence to have 93% sequence identity to SEQ ID NO:1. Claims 56, 59, and 64 require the nucleotide sequence to have 94% sequence identity to SEQ ID NO:1. Claims 38, 43, and 49 require the nucleotide sequence to have 95% sequence identity to SEQ ID NO:1.

Applicants submit that the disclosure in the application and the examples provided in the application and set forth in Table 1 are sufficient to meet the written description requirement for all of the claims. However, should the Board disagree, claims requiring 93%, 94%, and 95%

sequence identity are narrower in scope and thus should be considered separately in view of the arguments presented above.

9. ***Claims Appendix.***

An appendix containing a copy of the claims involved in the appeal.

10. ***Evidence Appendix.***

An appendix containing copies of the evidence submitted as follows:

Evidence Appendix A – Table 1

Evidence Appendix B – Lazar *et al.* (1988)

Evidence Appendix C – Hill *et al.* (1998)

Evidence Appendix D – de Maagd *et al.* (1999)

Evidence Appendix E – Angsuthanasombat *et al.* (2001)

Evidence Appendix F – Abad Declaration

Evidence Appendix G – Li *et al.* (1991)

CONCLUSION

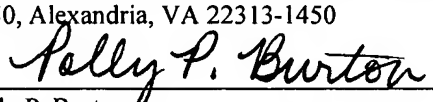
Appellants maintain that the Examiner has failed to carry her burden of establishing that the claims are not patentable because she has (a) failed to establish that it would require undue experimentation to practice the claimed invention and (b) failed to prove that the application does not adequately describe the claimed invention. Accordingly, claims 1-3, 9-12, 17-19, 38-40, 43-46, 49-52, 55-64 are allowable. For these reasons, presented in detail above, Appellants respectfully request that the rejections be reversed.

It is not believed that extensions of time are required. However, in the event that extensions of time are necessary to allow consideration of this paper, such extensions are hereby

petitioned under 37 CFR § 1.136(a), and any fee required therefore is hereby authorized to be charged to Deposit Account No. 16-0605.

Respectfully submitted,

W. Murray Spruill
Registration No. 32,943

CUSTOMER NO. 29122 **ALSTON & BIRD LLP** Bank of America Plaza 101 South Tryon Street, Suite 4000 Charlotte, NC 28280-4000 Tel Raleigh Office (919) 862-2200 Fax Raleigh Office (919) 862-2260	"Express Mail" mailing label number EV913516058 US Date of Deposit: October 10, 2006 I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to: Mail Stop Appeal Brief-Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450  Polly P. Burton

APPEALED CLAIMS

1. (previously presented) An isolated nucleic acid comprising a nucleotide sequence having at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1, wherein said nucleotide sequence encodes a polypeptide which is pesticidal for at least one pest belonging to the order Coleoptera.

2. (original) The nucleic acid according to claim 1, wherein the nucleotide sequence is optimized for expression in a plant.

3. (original) An expression cassette comprising a nucleic acid according to claim 1, wherein said nucleotide sequence is operably linked to a promoter that drives expression in a microorganism or in a plant cell.

4-8. (canceled)

9. (previously presented) A transformed plant comprising in its genome at least one stably incorporated nucleotide construct comprising a nucleotide sequence encoding a polypeptide operably linked to a promoter that drives expression of said polypeptide, wherein said polypeptide is pesticidal for at least one pest belonging to the order Coleoptera and wherein said nucleotide sequence has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1.

10. (original) The plant according to claim 9, wherein the plant is a monocot.

11. (original) The plant according to claim 9, wherein said plant is a dicot.

12. (original) Transformed seed of the plant according to claim 9.

13-16. (canceled)

17. (previously presented) A method for impacting an insect pest comprising introducing into a plant or cell thereof at least one nucleotide construct comprising a nucleotide sequence encoding a polypeptide operably linked to a promoter that drives expression of said polypeptide in plant cells, wherein said polypeptide is pesticidal for at least one pest belonging to the order Coleoptera and wherein said nucleotide sequence has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1, whereby an insect pest feeding on said plant or cell thereof is impacted.

18. (previously presented) The method according to claim 17, wherein the plant produces a polypeptide characterized by pesticidal activity against at least one pest of the order Coleoptera.

19. (previously presented) The method according to claim 18, wherein said insect pest is selected from the group consisting of Colorado potato beetle, western corn rootworm, southern corn rootworm, and boll weevil.

20-37. (canceled)

38. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence has at least 95% identity to the nucleotide sequence set forth in SEQ ID NO:1.

39. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:2.

40. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence is set forth in SEQ ID NO:1.

41-42. (canceled)

43. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence has at least 95% identity to the nucleotide sequence set forth in SEQ ID NO:1.

44. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:2.

45. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence is set forth in SEQ ID NO:1.

46. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence is optimized for expression in a plant.

47-48. (canceled)

49. (previously presented) The method of claim 17, wherein said nucleotide sequence has at least 95% identity to the nucleotide sequence set forth in SEQ ID NO:1.

50. (previously presented) The method of claim 17, wherein said nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:2.

51. (previously presented) The method of claim 17, wherein said nucleotide sequence is set forth in SEQ ID NO:1.

52. (previously presented) The method of claim 17, wherein said nucleotide sequence is optimized for expression in a plant.

53-54. (canceled)

55. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence has at least 93% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1.

56. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence has at least 94% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1.

57. (previously presented) The nucleic acid of claim 1, wherein said nucleotide sequence is optimized for expression in a plant.

58. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence has at least 93% sequence identity to the sequence set forth in SEQ ID NO:1.

59. (previously presented) The transformed plant of claim 9, wherein said nucleotide sequence has at least 94% sequence identity to the sequence set forth in SEQ ID NO:1.

60. (previously presented) The plant of claim 57, wherein said plant is a dicot.

61. (previously presented) The plant of claim 57, wherein said plant is a monocot.

62. (previously presented) The plant of claim 57, wherein said monocot is maize.

63. (previously presented) The method of claim 17, wherein said nucleotide sequence has at least 93% sequence identity to the sequence set forth in SEQ ID NO:1.

64. (previously presented) The method of claim 17, wherein said nucleotide sequence has at least 94% sequence identity to the sequence set forth in SEQ ID NO:1.

65. (canceled)

Table 1

SEQ ID NO for nucleotide sequence	Activity	Overall (global) % identity to SEQ ID NO:1	local % identity to nt 1-2007 of SEQ ID NO:1	Support in specification for activity	Corresponding polypeptide SEQ ID NO	Overall (global) % identity to SEQ ID NO:2	local % identity to aa 1-699 of SEQ ID NO:2
15	Activity against Colorado Potato Beetle	55%	100%	Table 1, page 68	16	56%	100%
19	Activity against Colorado Potato Beetle	52%	92.3%	Table 1, pp. 68-69	20	52%	92%
11	Activity against Colorado Potato Beetle, Southern Corn Rootworm	56%	99.4%	Table 1, pp. 68-69; Tables 2-4, pp. 70-71	12	56%	99.4%
23	Activity against Southern Corn Rootworm	56%	99.6%	Tables 2-4, pp. 70-71	24	56%	99.6%
31	Activity against Southern Corn Rootworm	51%	91.4%	Tables 2-4, pp. 70-71	32	51%	91.5%
33	Activity against Southern Corn Rootworm	51%	91.6%	Tables 2-4, pp. 70-71	34	60%	91.6%
29	Activity against Southern Corn Rootworm	51%	91.4%	Tables 2-4, pp. 70-71	30	51%	91.5%
9	Activity against Colorado Potato Beetle	38%	68.1%	Table 1, page 68	10	56%	100%

Table 1 (continued)

SEQ ID NO for nucleotide sequence	Activity	Overall (global) % identity to SEQ ID NO:1	local % identity to nt 1-2007 of SEQ ID NO:1	Support in specification for activity	Corresponding polypeptide SEQ ID NO	Overall (global) % identity to SEQ ID NO:2	local % identity to aa 1-699 of SEQ ID NO:2
43	Activity against Coleopteran Pests, data not shown	56%	99.6%	Page 27 of specification, lines 22-25	44	56%	99.6%
21	Activity against Coleopteran Pests, data not shown	56%	99.4%	Page 27 of specification, lines 22-25	22	56%	99.4%
39	Activity against Coleopteran Pests, data not shown	56%	99.4%	Page 27 of specification, lines 22-25	40	56%	99.4%
41	Activity against Coleopteran Pests, data not shown	51%	91.3%	Page 27 of specification, lines 22-25	42	51%	91.5%
45	Activity against Coleopteran Pests, data not shown	51%	91.6%	Page 27 of specification, lines 22-25	46	51%	91.6%
3	Activity against Coleopteran Pests, data not shown	92%	NA	Page 11, lines 15-19. examples 3 and 4	4	89%	NA

MOLECULAR AND CELLULAR BIOLOGY, Mar. 1988, p. 1247–1252
0270-7306/88/031247-06$02.00/0

Vol. 8, No. 3

Transforming Growth Factor α: Mutation of Aspartic Acid 47 and Leucine 48 Results in Different Biological Activities

ELIANE LAZAR,† SHINICHI WATANABE,* STEPHEN DALTON, AND MICHAEL B. SPORN

Laboratory of Chemoprevention, National Cancer Institute, Bethesda, Maryland 20892

Received 22 July 1987/Accepted 30 November 1987

To study the relationship between the primary structure of transforming growth factor α (TGF-α) and some of its functional properties (competition with epidermal growth factor (EGF) for binding to the EGF receptor and induction of anchorage-independent growth), we introduced single amino acid mutations into the sequence for the fully processed, 50-amino-acid human TGF-α. The wild-type and mutant proteins were expressed in a vector by using a yeast α mating pheromone promoter. Mutations of two amino acids that are conserved in the family of the EGF-like peptides and are located in the carboxy-terminal part of TGF-α resulted in different biological effects. When aspartic acid 47 was mutated to alanine or asparagine, biological activity was retained; in contrast, substitutions of this residue with serine or glutamic acid generated mutants with reduced binding and colony-forming capacities. When leucine 48 was mutated to alanine, a complete loss of binding and colony-forming abilities resulted; mutation of leucine 48 to isoleucine or methionine resulted in very low activities. Our data suggest that these two adjacent conserved amino acids in positions 47 and 48 play different roles in defining the structure and/or biological activity of TGF-α and that the carboxy terminus of TGF-α is involved in interactions with cellular TGF-α receptors. The side chain of leucine 48 appears to be crucial either indirectly in determining the biologically active conformation of TGF-α or directly in the molecular recognition of TGF-α by its receptor.

Transforming growth factor α (TGF-α) is a polypeptide of 50 amino acids. First isolated from a retrovirus-transformed mouse cell line (9), it has subsequently been found in human tumor cells (10, 29), in the early rat embryo (18), and recently in cell cultures from the pituitary gland (23). TGF-α appears to be closely related to epidermal growth factor (EGF) structurally and functionally (19, 20). The two peptides apparently bind to the same receptor, and both induce anchorage-independent growth of certain nontransformed cells, such as NRK cells, in the presence of TGF-β (1).

Comparison of amino acid sequences reveals about 35% homology among the EGF-like peptides (rat [27], mouse [25], and human [13] EGFs and rat [19] and human [12] TGF-αs). Some viral peptides (Shope fibroma growth factor [6], vaccinia growth factor [2], and myxoma growth factor [30]) also share homologies with the EGF-like peptides.

If TGF-α is involved in transformation, a TGF-α antagonist could be an important therapeutic tool in the treatment of certain types of malignancies. An understanding of the conformational and dynamic properties of the TGF-α molecule is basic to the design of an antagonist. A hypothetical antagonist would bind to the same receptor as TGF-α, but would not induce the series of proliferative and transforming events induced by TGF-α. To obtain such a molecule it is necessary to dissociate interactions responsible for binding from those involved in signal transduction. We decided to approach the problem by way of site-directed mutagenesis of a human sequence of TGF-α. In this report we describe our first series of mutations, which were carried out at residues Asp-47 and Leu-48, in the carboxy-terminal part of TGF-α; these two amino acids are highly conserved in the EGF-like family of peptides. We show that these two adjacent residues play different roles in the structure and/or function of TGF-α.

* Corresponding author.
† Present address: Unité d'Oncologie Moleculaire, IRSC, 94802 Villejuif Cedex, France.

MATERIALS AND METHODS

Cells. Normal rat kidney (NRK) cells were grown in Dulbecco modified Eagle medium containing 10% (vol/vol) calf serum.

TGF-α gene. The sequence of the 50-amino-acid human TGF-α was originally derived from a human TGF-α precursor cDNA (12). The coding sequence is preceded by an ATG methionine codon and followed by a TAA stop codon and is flanked by *Eco*RI restriction sites. This *Eco*RI fragment combines the 59-base-pair *Eco*RI-*Nco*I fragment from plasmid pTE5 (12) with the 111-base-pair *Nco*I-*Eco*RI fragment from plasmid pyTE2 (11). The resulting *Eco*RI fragment was inserted in M13mp18 for site-directed mutagenesis.

Synthesis and purification of oligonucleotides and oligonucleotide-directed mutagenesis. The synthesis and purification of 20- to 27-nucleotide oligonucleotides were carried out as described previously (31). The one or two nucleotides responsible for the mutation were located in the middle of the oligonucleotide. Mutagenesis was performed by published procedures (21, 33). The sequences of the mutant clones were verified by the method of Sanger et al. (25).

Yeast shuttle vector. The vector YEp70αT contains a yeast α-factor pheromone promoter and prepro sequence for the expression of TGF-α (15). The mutant TGF-α coding sequence was inserted in the *Eco*RI site of plasmid YEp70αT and expressed in the form of a fusion protein consisting of 92 amino acids from the prepro sequence of the yeast α factor attached to the amino terminus of TGF-α (28). The yeast cleaves the precursor and secretes TGF-α with 8 amino acids fused to it (4 are encoded by the prepro sequence of α-factor, and the other 4 are encoded by the DNA sequence added to insert of the TGF-α gene). The last of these residues is a methionine, which allows the cleavage of the secreted fusion

protein by cyanogen bromide (CNBr) and the release of a mature TGF-α (50 amino acids) (see Results).

Yeast strain and transformation. The yeast *Saccharomyces cerevisiae* 20B-12 (*MATα trp1 pep4-3*) (17) was obtained from the Yeast Genetics Stock Center, Berkeley, Calif. *S. cerevisiae* 20B-12 was grown in YEPD medium (1% yeast extract [Difco Laboratories], 2% Bacto-Peptone [Difco], 2% glucose). When the culture reached an optical density at 660 nm of 1, spheroplasts were prepared (14) for transformation. For each transformation we used 10 to 15 μg of purified plasmid DNA.

Partial purification of TGF-α mutants. At 3 days after transformation, five individual colonies of transformants were grown to saturation in YEPD medium. The amount of protein in the yeast medium was measured by the method of Bradford (3), and the amount of mutant TGF-α secreted in the yeast medium was determined by radioimmunoassay. The clones which secrete the highest amount of mutant TGF-α were used to grow a 1-liter culture in YNB-CAA medium (0.67% yeast nitrogen base, 20 g of glucose per liter, 10 g of Casamino Acids [Difco] per liter). After the culture reached saturation (optical density at 660 nm of 10 to 12) (48 h in an air shaker at 30°C), the yeast conditioned medium was dialyzed extensively against 1 M acetic acid in 3,000-molecular-weight cutoff dialysis tubing. Usually 250 ml of dialyzed culture was lyophilized, suspended in 10 ml of 70% formic acid, and treated with CNBr (molar excess of 500) for 20 h at room temperature. The CNBr was subsequently evaporated, and the samples were lyophilized. CNBr-treated samples were suspended in 1 ml of 1 M acetic acid, loaded on a Bio-gel P30 column (30 by 1.5 cm [Bio-Rad Laboratories]), and eluted with 1 M acetic acid. Fractions of 1 ml were collected. Aliquots were lyophilized, suspended in binding buffer (minimum essential medium containing 1 mg of bovine serum albumin per ml and 25 mM HEPES [*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid; pH 7.4]), neutralized if necessary to pH 7.4, and tested in EGF-binding competition and soft-agar assays, as well in radioimmunoassay.

Radioimmunoassays. The amounts of TGF-α secreted in the yeast medium were determined by radioimmunoassay with the immunoglobulin G fraction of a polyclonal antibody, 34D, raised against recombinant human TGF-α (4), in 0.1 M Tris (pH 7.5)–0.15 M NaCl–2.5 mg of bovine serum albumin per ml. The amounts of partially purified TGF-α present in the P30 column fractions were measured by using the Biotope RIA kit with polyclonal antibody against human TGF-α (a gift from W. Hargreaves, Biotope), under denaturing conditions, as recommended by the supplier.

EGF binding competition assay and soft agar assay. Both EGF-binding competition and soft-agar assays have been described previously (1).

RESULTS

Rationale for mutations in the carboxyl terminus of TGF-α. Figure 1 shows the amino acid sequence of TGF-α in which the residues that are conserved among all the EGF-like peptides described thus far (EGF, TGF-α, and EGF-like viral proteins) are enclosed in bold circles. Among the 11 conserved amino acids, there are 6 Cys and 2 Gly residues, which presumably play essential roles in determining the overall conformation of the molecule. We concentrated on the two conserved amino acids in the carboxyl terminus, Asp-47 and Leu-48. The Asp in position 47 is conserved among the EGFs and TGF-α (human or murine), but not

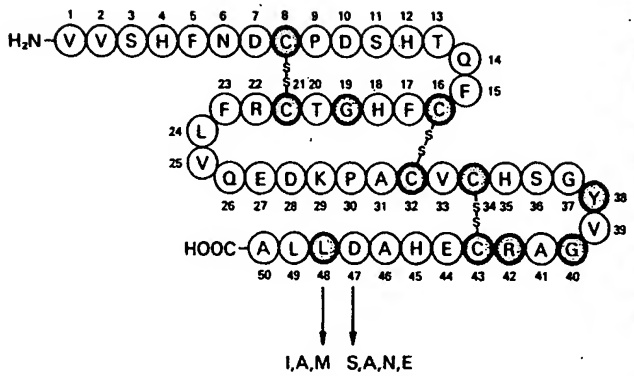


FIG. 1. Mutations in the carboxy terminus of human TGF-α. The amino acids conserved in all the family of EGF-like growth factors (human and murine EGFs and TGFs, as well as the gene products of the vaccinia virus [vaccinia growth factor], the Shope fibroma virus [Shope fibroma growth factor], and the myxoma virus [myxoma growth factor]) are enclosed in bold circles. The mutations of amino acids at positions 47 and 48 are indicated. Symbols: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

among the EGF-like viral proteins (vaccinia growth factor, Shope fibroma growth factor, or myxoma growth factor), whereas Leu 48 is conserved among all the EGF-like peptides so far described. In both mouse and human EGF, the two corresponding residues (Asp-46 and Leu-47) are located near the surface of the protein (8, 22, 22a). We designed a series of mutations in these two positions.

Asp-47 has been mutated to Glu, Asn, Ser, and Ala. Glu was chosen because it has the same charge as and a larger size than Asp; Asn has a similar side-chain structure, but is uncharged; Ser is smaller but still polar; Ala is smaller and nonpolar.

Leu 48 has been mutated to Ile and Met, which are both large, nonpolar residues like Leu, and to Ala, which is nonpolar but smaller. We introduced the chosen mutations by site-directed mutagenesis of the cloned human TGF-α gene, using synthetic oligonucleotides.

Construction of the yeast α mating pheromone–human TGF-α plasmid. The TGF-α expression vector pyTE1 (Fig. 2) was constructed by using plasmid YEp70αT (15) which contains the 2μm origin of replication and yeast *TRP1* gene for its replication and selective maintenance, respectively. YEp70αT also contains the yeast α-factor promoter, the α-factor prepro sequence coding for 89 amino acids, and the sequence for 3 amino acids resulting from the introduction of *Xba*I and *Eco*RI sites. The human mature TGF-α sequence (12) is contained in a 170-base-pair *Eco*RI fragment which includes an ATG (Met) codon preceding the sequence of TGF-α and a TAA (stop) codon followed by 8 nucleotides. This TGF-α sequence was inserted in the unique *Eco*RI site of YEp70αT. Clones with the proper orientation were selected, and DNA was isolated for yeast transformation.

Measurement of TGF-α secreted by *S. cerevisiae*. The amount of total proteins secreted into the yeast culture was 10 ± 1 μg/ml for wild-type as well as mutant TGF-α as determined by the method of Bradford (3). Before further purification was attempted, we wanted to determine whether the mutated TGF-α proteins were being secreted by the yeast. The low pH of the yeast medium, as well as the acidic proteins secreted in the yeast culture, precluded biological assay of secreted mutants. Therefore, immunological meth-

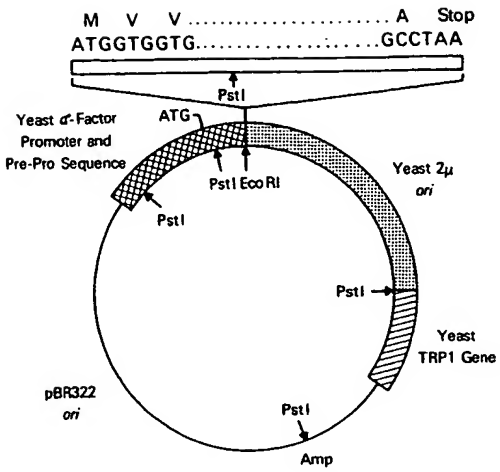


FIG. 2. Structure of the *S. cerevisiae* 8.2-kilobase shuttle vector pyTE1. The secretion of the TGF-α gene is under the transcriptional control of the yeast α-factor promoter and prepro sequence (▩). The yeast 2μm origin of replication (▨) and the selective yeast *TRP1* gene (▧) are indicated. The TGF-α gene, preceded by an initiation (ATG) codon and followed by a stop (TAA) codon, is inserted in the *Eco*RI site. Details are given in Materials and Methods and in Results.

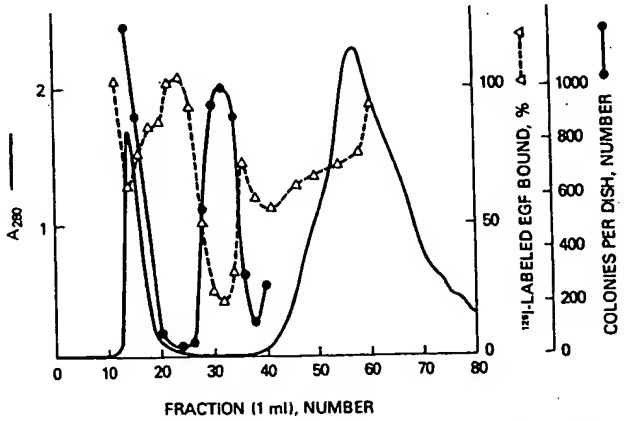


FIG. 3. Purification of yeast-secreted wild-type TGF-α. The purification procedure is described in Materials and Methods and in Results. Aliquots of every other fraction of the Bio-Gel P30 column were tested for their abilities to compete with ^{125}I-EGF for binding to the EGF receptor (Δ) and to induce colony formation (>62 μm) on NRK cells in soft agar in the presence of TGF-β (1 ng/ml) (●). The A_{280} profile of the proteins was determined (——).

ods were used. Wild-type and mutant TGF-α's were secreted at a level of 100 to 200 ng/ml and 10 to 500 ng/ml, respectively (as determined by radioimmunoassay with polyclonal antibody 34D). We thus estimate that the percentage of TGF-α secreted in the yeast culture is at least 1% of the total protein secreted. We cannot yet assess whether the variations in the levels of secretion of different mutant TGF-α proteins are real or whether one single-amino-acid substitution drastically affects the recognition by the antibody. The latter hypothesis is the more likely, since the use of another polyclonal antibody (Biotope) under denaturing conditions enabled us to detect certain TGF-α mutants (such as [Ala 47]–TGF-α, in which the amino acid in position 47 of human TGF-α is mutated to an alanine) that were poorly detected by 34D, under nondenaturing as well as denaturing conditions. After the amount of TGF-α mutant proteins was estimated, the medium was extensively dialyzed against 1 M acetic acid and lyophilized as described in Materials and Methods.

Partial purification of yeast-secreted TGF-α. Although the yeast shuttle vector was constructed in such a way as to secrete TGF-α with 8 amino acids fused to the N terminus, it was often observed that a significant fraction of the secreted TGF-α was in a higher-molecular-weight fragment corresponding to the size expected from an uncleaved (unprocessed) 92-amino-acid fusion protein. Since a Met had been introduced at the N terminus of TGF-α and since TGF-α contains no Met in its sequence, CNBr treatment could be used to cleave either of these 8- or 92-amino-acid N-terminal peptides and release the complete 50-amino-acid TGF-α. Indeed, CNBr treatment of yeast-secreted proteins resulted in the conversion of high-molecular-weight TGF-α into the 6,000-molecular-weight species, as revealed by Western immunoblot (data not shown).

CNBr-cleaved samples (see Materials and Methods) were purified on a Bio-Gel P30 column. Figure 3 shows the elution profile of the proteins, as well as the results of a radioreceptor assay and a soft-agar assay performed on aliquots of the column fractions. The A_{280} profile shows two major peaks of eluted proteins, one corresponding to the void volume and the other one to proteins of molecular weight <3,000. Aliquots of the column fractions were tested for their ability to compete with ^{125}I-EGF for binding to the receptor. The fractions that were the most active in this assay were located between the two major protein peaks, in an area where relatively few proteins eluted. Although some activity was found in the first protein peak (void volume), this was considerably reduced on treatment with stronger CNBr (data not shown).

Aliquots of each fraction were also tested for their ability to induce anchorage-independent growth of NRK cells in soft agar in the presence of TGF-β (1 ng/ml). The receptor binding and colony-forming activity superimposed almost exactly (Fig. 3). Analysis by polyacrylamide gel electrophoresis with silver staining, as well as by Western blot, of the column fractions shows that our purification procedure (CNBr cleavage followed by P30 sizing column) eliminates high-molecular-weight proteins (data not shown). Since pure TGF-α migrates in a broad band on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (32), this technique cannot be used for proper assessment of the degree of separation of TGF-α from low-molecular-weight contaminating proteins. Nevertheless, within our detection levels the amounts of TGF-α present in the column fractions (detected by radioimmunoassay using the antibody from Biotope) correlated with the amounts observed on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (data not shown).

Comparison of binding and colony-forming activity of TGF-α partially purified from yeast media. It was important to show that wild-type TGF-α secreted from *S. cerevisiae* had the expected biological properties and that its activity in soft-agar and radioreceptor assays was equivalent. For these assays, the amount of EGF-competing activity present in the most active fraction of the P30 column of wild-type TGF-α was measured in terms of EGF equivalents. The dilution curve had a slope that was parallel to that of the EGF standard. This value was also used to measure the colony-forming activity of the partially purified wild-type TGF-α (with EGF as a standard in the assay). The colony-forming activity of the partially purified wild-type TGF-α corre-

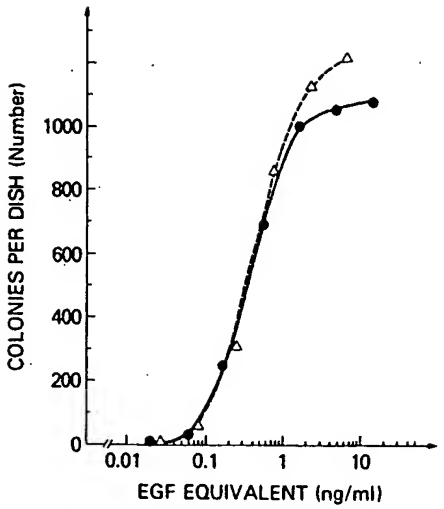


FIG. 4. Correlation between the activities in the binding and colony-forming assay for the partially purified wild-type TGF-α secreted by *S. cerevisiae*. The activity in the radioreceptor assay of the peak fraction from the P30 column was determined in EGF equivalent concentration. The value obtained was used for the soft-agar assay. Colonies of >62 μm (△) and the EGF standard (●) are shown.

sponded exactly to that of EGF (Fig. 4). Thus, we have partially purified a wild-type 50-amino-acid TGF-α showing the expected binding and colony-forming activities, which provides a reference substance for mutant TGF-αs that might show a dissociation of binding and colony-forming abilities.

Biological and biochemical activities of the partially purified TGF-α mutant proteins. Mutated TGF-αs were expressed by using the yeast system and partially purified on Bio-Gel P30 columns as described in Materials and Methods. Mutant TGF-αs were usually obtained from two different clones of yeast transformants. The CNBr-cleaved samples were purified through different Bio-Gel P30 columns for each mutant protein to avoid any possible contamination from one peptide to another. The purification profiles observed with the mutant TGF-αs were similar to those obtained for the wild-type TGF-α. Aliquots of the P30 column fractions were tested in radioreceptor and soft-agar assays. For all mutant proteins, the highest activity in both assays was always found in the same fraction of the Bio-Gel P30 column effluent (peak fraction). Extensive purification of a series of mutant proteins for screening purposes is not practical. Therefore, we needed a quantitation system that would allow us to compare mutant proteins with each other. Thus, the amount of TGF-α present in the peak fraction was estimated by radioimmunoassay with an antiserum to native TGF-α (obtained from W. Hargreaves), under denaturing conditions, as described in Materials and Methods. All values given in Table 1 were obtained from the peak fraction.

The controls done with the wild-type TGF-α showed (Fig. 4; Table 1) that binding and transforming activity were equivalent. The yeast vector without a TGF-α insert did not secrete any EGF-like proteins, as determined by both radioreceptor and soft-agar assay.

Two types of results were obtained upon assay of mutant proteins having different amino acid substitutions at Asp-47. In both [Ala-47]–TGF-α and [Asn-47]–TGF-α, binding ability was retained. Soft-agar and radioreceptor activities correlated for [Asn-47]–TGF-α; there was a lower value for

TABLE 1. Biological and biochemical activities of mutant TGF-α proteins secreted by *S. cerevisiae* and partially purified

Insert in the yeast expression vector	EGF equivalence (ng/ml) in:		Amt of TGF-α (ng/ml) in radioimmunoassay
	Radioreceptor assay	Soft-agar assay	
Wild-type TGF-α	700	700	2,000
	400	300	ND[a]
None	0	0	0
[Ala-47]–TGF-α	100	44	220
	66	48	ND
[Asn-47]–TGF-α	80	72	180
	75	72	525
[Glu-47]–TGF-α	3	3	42
[Ser-47]–TGF-α	10	4	60
[Ala-48]–TGF-α	0	0	16
	0	0	220
[Ile-48]–TGF-α	4	12	470
	2	7	490
[Met-48]–TGF-α	2	8	453
	0.5	2	420

[a] ND, Not determined.

colony-forming activity than for EGF-binding competition for [Ala-47]–TGF-α. [Ser-47]–TGF-α and [Glu-47]–TGF-α appeared to have lower activities in both assays than either wild-type TGF-α or [Ala-47]–TGF-α and [Asn-47]–TGF-α. These results indicate that neither the carboxyl charge nor the polarity of Asp-47 is essential for biological activity.

The effects of mutation of Leu-48, one of the 11 amino acids perfectly conserved among all the EGFs, TGF-αs, and viral EGF-like proteins, are dramatic. [Ala-48]–TGF-α totally lacked binding and colony-forming activity. [Ile-48]–TGF-α and [Met-48]–TGF-α had very little biological activity compared with wild-type TGF-α. Another substitution, [Met-48]–TGF-α, resulted in a truncated mutant lacking the last 2 amino acids and having a substitution of Leu to homoserine at position 48 following treatment with CNBr. Alternatively, if [Met-48]–TGF-α was not treated with CNBr, fusion proteins of TGF-α (mutated to Met in position 48) with 8 or 92 amino acids attached at the N terminus were obtained. Very low activities in binding and soft-agar assays were found for these mutants, whether or not they were cleaved with CNBr. Experiments on EGF and TGF-α have shown that an N-terminal extension does not markedly modify EGF-binding activity (12, 26). Therefore, the loss of activity obtained with [Met-48]–TGF-α that has not been CNBr treated was probably due to the mutation itself and not to the N-terminally extended fusion protein. We do not know whether the loss of activity observed with the TGF-α shortened to 48 amino acids and having a substitution of Leu-48 to homoserine is due only to the mutation or also to the lack of the last 2 amino acids.

The data obtained by radioimmunoassay on the partially purified wild-type and mutant TGF-α show that the amount of TGF-α detected was always higher than the amount determined by measurement of biological activity. This may be due to the presence in the fraction of a certain percentage of incorrectly folded TGF-α that might be recognized in a

radioimmunoassay under denaturing conditions but would not be biologically active. None of the mutant proteins seemed to be present in amounts equivalent to those observed for wild-type TGF-α in the partially purified fractions (whether radioimmunoassay, radioreceptor, or soft-agar assay was used for quantitation). It is not clear whether consistently less TGF-α was produced by the mutant constructs than by the wild type or whether the secreted mutant proteins were simply less well recognized by the antibody. Because of these uncertainties, the biological activities of the different mutant proteins cannot be accurately related to a known amount of mutant TGF-α protein. Even though radioimmunoassay should be used with caution for a quantitative evaluation of mutant TGF-α proteins, a positive reaction demonstrates that immunoreactive TGF-α was present in the P30 peak fraction for each mutant. Therefore, the fact that one of the mutant proteins ([Ala-48]–TGF-α) is biologically inactive can be attributed to the mutation itself, and not to the lack of production of the mutant protein by the yeast or its loss through purification. However, if the mutant proteins are in fact as immunoreactive as the wild type, then [Ala-47]–TGF-α and [Asn-47]–TGF-α are as active as wild-type TGF-α and [Glu-47]–TGF-α and [Ser-47]–TGF-α are less active; in contrast, [Ile-48]–TGF-α and [Met-48]–TGF-α are almost inactive. The differences between mutation of Asp-47 and Leu-48 would then be even more striking.

DISCUSSION

TGF-α shows sequence homologies with EGF, and both growth factors share the same cellular receptors (20). Even though EGF was discovered 25 years ago (7) and its properties have been extensively studied over the years (5), the binding site of EGF to its receptor has still not been determined, and the relationship between structure and function of EGF/TGF-α is still to be discovered. Particularly, we do not know whether binding to the receptor and signal transduction occur through one or more domains of the molecule or through which amino acids. We approached the question by performing site-directed mutagenesis of TGF-α and focused our attention on two adjacent amino acids, Asp-47 and Leu-48, located in the carboxy terminus and highly conserved in the EGF-like family of peptides. Unexpectedly, these two amino acids showed very different sensitivities to mutation and particularly to a substitution to Ala: [Ala-47]–TGF-α retained binding and colony-forming activities, whereas [Ala-48]–TGF-α completely lost both activities. These data show that Asp-47 and Leu-48 play very different roles in defining the structure and/or the activity of TGF-α. The other mutations performed on Asp-47 were substitutions to Asn, Ser, and Glu. [Asn-47]–TGF-α, like [Ala-47]–TGF-α, was active in binding and induction of colony formation, but [Ser-47]–TGF-α and [Glu-47]–TGF-α showed weaker growth factor activities. These results indicate that neither the carboxyl charge nor the polarity of Asp-47 is essential for biological activity. Interestingly, two of the EGF-like viral proteins, myxoma growth factor and Shope fibroma growth factor (6, 30), have Asn instead of Asp in position 47; we have shown that [Asn-47]–TGF-α retains biological activity.

Substitution of Leu-48 to Met and Ile led to mutant proteins with very low activities, whereas substitution to Ala led to complete loss of activity. We did not expect that a mutation of Leu to Ile (which have similar sizes and polarities) would cause such a strong effect. Thus, Leu-48, which is conserved perfectly among all the EGF-like peptides, seems to be essential, through its exact geometry, for the biological activity of TGF-α.

The mutant proteins tested so far, when active, showed parallel behaviors in binding and colony formation. Some mutant proteins lost all activities, and we assume that the binding capacity has been lost. We have not been able to dissociate the binding and colony-forming abilities by using any of the present series of mutant proteins, and it is necessary to screen more of them in search of an antagonist of TGF-α.

Results relating to the biological activity of EGF show that derivatives of mouse EGF and human EGF (EGF 1-47) lacking the carboxy-terminal 6 amino acids as a result of enzymatic digestion are less potent than the intact molecule in mitogenic stimulation of fibroblasts, but retain full biological activity in in vivo assays (inhibition of gastric acid secretion) (16). On the other hand, naturally occurring truncated forms of rat EGF, which lack the carboxy-terminal 5 amino acids (rEGF 2-48) are as potent as mouse EGF (mEGF 1-53) in receptor-binding and mitogenic assays (27). We do not know whether the discrepancies observed are due to the origin of the molecule (artificial or natural) or to the type of bioassay used. In any event, all of these EGF-related molecules, which are shorter than mouse or human EGF, still retain Leu-47. We have shown that in TGF-α, the corresponding residue, Leu-48, is critical for the biological activity.

Recent data on the three-dimensional structure of mouse EGF obtained by nuclear magnetic resonance show that even though Asp-46 and Leu-47 (Asp-47 and Leu-48 in TGF-α) are both solvent accessible (8, 22, 22a), their side chains point in opposite directions in the beta-sheet structure. Therefore, the role of these adjacent amino acids in the structure and, consequently, the function of EGF might be very different. Our data show that the amino acids Asp-47 and Leu-48 of TGF-α are not equally important for the biological activity of TGF-α, despite their conservation among the EGF-like peptides. From the dramatic loss in biological activity which is characteristic of mutation of Leu-48, we also suggest that this residue is involved in binding to the cellular receptors either by direct interaction with the receptor or by providing the proper conformation to the molecule.

ACKNOWLEDGMENTS

We thank Rik Derynck (Genentech) for providing the TGF-α gene, inserted in M13, and for his assistance throughout this project. We are indebted to Arjun Singh (Genentech) for helping us with the yeast transformation and expression. We are grateful to Tim Bringman (Genentech) and William Hargreaves (Biotope) for their generous gifts of TGF-α antibodies. We thank Linda Durham for technical assistance, Irene Dalton for manuscript preparation, and our colleagues for helpful comments and moral support.

LITERATURE CITED

1. **Anzano, M. A., A. B. Roberts, J. M. Smith, M. B. Sporn, and J. E. De Larco.** 1983. Sarcoma growth factor from conditioned medium of virally transformed cells is composed of both type α and type β transforming growth factors. Proc. Natl. Acad. Sci. USA 80:6264–6268.
2. **Blomquist, M. C., L. T. Hunt, and W. C. Barker.** 1984. Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. Proc. Natl. Acad. Sci. USA 81:7363–7367.
3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254.

4. **Bringman, T. S., P. B. Lindquist, and R. Derynck.** 1987. Different transforming growth factor-α species are derived from a glycosylated and palmitoylated transmembrane precursor. Cell **48**:429–440.
5. **Carpenter, G., and S. Cohen.** 1979. Epidermal growth factor. Annu. Rev. Biochem. **48**:193–216.
6. **Chang, W., C. Upton, S. Hu, A. F. Purchio, and G. McFadden.** 1987. The genome of Shope fibroma virus, a tumorigenic poxvirus, contains a growth factor gene with sequence similarity to those encoding epidermal growth factor and transforming growth factor α. Mol. Cell. Biol. **7**:535–540.
7. **Cohen, S.** 1962. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. J. Biol. Chem. **237**:1555–1562.
8. **Cooke, R. M., A. J. Wilkinson, M. Baron, A. Pastore, M. J. Tappin, I. D. Campbell, H. Gregory, and B. Sheard.** 1987. The solution structure of human epidermal growth factor. Nature (London) **327**:339–341.
9. **De Larco, J. E., and G. J. Todaro.** 1978. Growth factors from murine sarcoma virus-transformed cells. Proc. Natl. Acad. Sci. USA **75**:4001–4005.
10. **Derynck, R., D. V. Goeddel, A. Ullrich, J. U. Gutterman, R. D. Williams, T. S. Bringman, and W. H. Berger.** 1987. Synthesis of messenger RNAs for transforming growth factors α and β and the epidermal growth factor receptor by human tumors. Cancer Res. **47**:707–712.
11. **Derynck, R., A. B. Roberts, D. H. Eaton, M. E. Winkler, and D. V. Goeddel.** 1985. Human transforming growth factor-α: precursor sequence, gene structure, and heterologous expression. Cancer Cells **3**:79–86.
12. **Derynck, R., A. B. Roberts, M. E. Winkler, E. Y. Chen, and D. V. Goeddel.** 1984. Human transforming growth factor-α: Precursor structure and expression in E. coli. Cell **38**:287–297.
13. **Gregory, H.** 1975. Isolation and structure of urogastrone and its relationship to epidermal growth factor. Nature (London) **257**:325–327.
14. **Hinnen, A., J. B. Hicks, and G. R. Fink.** 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
15. **Hitzeman, R. A., C. N. Chang, M. Matteucci, L. J. Perry, W. J. Kohr, J. J. Wulf, J. R. Swartz, C. Y. Chen, and A. Singh.** 1986. Construction of expression vectors for secretion of human interferons by yeast. Methods Enzymol. **119**:424–433.
16. **Hollenberg, M. D., and H. Gregory.** 1980. Epidermal growth factor—urogastrone: biological activity and receptor binding of derivatives. Mol. Pharmacol. **17**:314–320.
17. **Jones, E.** 1976. Proteinase mutants of *Saccharomyces cerevisiae*. Genetics **85**:23–30.
18. **Lee, D. C., R. Rochford, G. J. Todaro, and L. P. Villarreal.** 1985. Developmental expression of rat transforming growth factor α mRNA. Mol. Cell. Biol. **5**:3644–3646.
19. **Marquardt, H., M. W. Hunkapiller, L. E. Hood, and G. J. Todaro.** 1984. Rat transforming growth factor type 1: structure and relation to epidermal growth factor. Science **223**:1079–1082.
20. **Massague, J.** 1983. Epidermal growth factor-like transforming growth factor. II. Interaction with epidermal growth factor receptors in human placenta membranes and A431 cells. J. Biol. Chem. **258**:13614–13620.
21. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
22. **Montelione, G. T., K. Wuthrich, E. C. Nice, A. W. Burgess, and H. A. Scheraga.** 1986. Identification of two anti-parallel beta-sheet conformations in the solution structure of murine epidermal growth factor by proton magnetic resonance. Proc. Natl. Acad. Sci. USA **83**:8594–8598.

22a. **Montelione, G. J., K. Wuthrich, E. C. Nice, A. W. Burgess, and H. A. Scheraga.** 1987. Solution structure of murine epidermal growth factor: determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. Proc. Natl. Acad. Sci. USA **84**:5226–5230.

23. **Samsoondar, J., M. S. Kobrin, and J. E. Kudlow.** 1986. α-Transforming growth factor secreted by untransformed bovine anterior pituitary cells in culture. J. Biol. Chem. **261**: 14408–14413.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Savage, C. R., Jr., J. H. Hash, and S. Cohen.** 1973. Epidermal growth factor: location of disulfide bonds. J. Biol. Chem. **248**:7669–7672.
26. **Shechter, Y., J. Schlessinger, S. Jacobs, K. J. Chang, and P. Cuatrecasas.** 1978. Fluorescent labeling of hormone receptors in viable cells: preparation and properties of highly fluorescent derivatives of epidermal growth factor and insulin. Proc. Natl. Acad. Sci. USA **75**:2135–2139.
27. **Simpson, R. J., J. A. Smith, R. L. Moritz, M. J. O'Hare, P. S. Rudland, J. R. Morrison, C. J. Lloyd, B. Grego, A. W. Burgess, and E. C. Nice.** 1985. Rat epidermal growth factor: complete amino acid sequence. Eur. J. Biochem. **153**:629–637.
28. **Singh, A., J. M. Lugovoy, W. J. Kohr, and L. J. Perry.** 1984. Synthesis, secretion and processing of α-factor-interferon fusion proteins in yeast. Nucleic Acids Res. **12**:8927–8938.
29. **Todaro, G. J., C. Fryling, and J. E. DeLarco.** 1980. Transforming growth factors produced by certain human tumors: polypeptides that interact with epidermal growth factor receptors. Proc. Natl. Acad. Sci. USA **77**:5258–5262.
30. **Upton, C., J. L. Macen, and G. McFadden.** 1987. Mapping and sequencing of a gene from myxoma virus that is related to those encoding epidermal growth factor and transforming growth factor α. J. Virol. **61**:1271–1275.
31. **Watanabe, S., E. Lazar, and M. B. Sporn.** 1987. Transformation of normal rat kidney (NRK) cells by an infectious retrovirus carrying a synthetic rat type α transforming growth factor gene. Proc. Natl. Acad. Sci. USA **84**:1258–1262.
32. **Winkler, M. E., T. Bringman, and B. J. Marks.** 1986. The purification of fully active recombinant transforming growth factor-α produced in *Escherichia coli*. J. Biol. Chem. **261**:13838–13843.
33. **Zoller, M. J., and M. Smith.** 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Methods Enzymol. **101**:468–500.

March No 2

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS 244, 573–577 (1998)
ARTICLE NO. RC988301

Functional Analysis of Conserved Histidines in ADP-Glucose Pyrophosphorylase from *Escherichia coli*[1]

Margaret A. Hill[2] and Jack Preiss[3]
Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Received February 11, 1998

Two absolutely conserved histidines and a third highly conserved histidine are noted in 11 bacterial and plant ADP-glucose pyrophosphorylases. These histidines were individually mutagenized in the *E. coli* enzyme to glutamine in order to determine their function. Glutamine mutations at residues 143 and 156 produced functional enzymes in cell extracts with slightly lower than wild-type specific catalytic activities and with same heat stability characteristics of the wild-type enzyme. Substitution of residue 83 with glutamine however produced an enzyme having decreased thermal stability. Additional mutageneses at residue 83 with asparagine, arginine, or aspartate gave rise to enzymes having a progressively decreasing trend in thermal stability. These mutants are more susceptible to proteolysis than wild-type enzyme. Kinetic analysis of H83Q and H83N indicates that histidine 83 is not involved in the catalytic mechanism or in substrate binding but possibly in maintenance of the active catalytic structure.

ADP-glucose pyrophosphorylase (ATP:a-glucose-1-phosphate adenylyltransferase, EC 2.7.7.27) is an example of an allosterically regulated enzyme whose general function is conserved in evolutionarily widely separated organisms. Bacterial and higher plant enzymes catalyze the reversible reaction between ATP and glucose 1-phosphate to yield ADP-glucose and pyrophosphate. Metabolically, this represents the first step in glycogen and starch biosynthesis in bacteria and plants, respectively. Allosteric regulation by glucose metabolites modulates enzyme activity; however, the structural specificity of the allosteric modifiers is different for the bacterial and plant enzymes, reflecting their differing assimilating glucose metabolic pathways.

Details of the mechanisms of catalysis and regulation for this enzyme provide insight into general principles of protein evolution of structure. Although tertiary structure information is presently unavailable for this enzyme, primary sequence data has been collected from a large array of bacterial, cyanobacterial, and higher plant pyrophosphorylases (1). Comparisons of these sequences highlight those strictly conserved residues whose functions are essential. Elucidation of their specific roles in enzyme function have been used by site-directed mutagenesis techniques. This study initiates studies on the structure-function roles of histidine residues for the *E.coli* ADP-glucose pyrophosphorylase.

Histidines were chosen as targets for mutagenesis studies due to their potential role in catalytic mechanisms involving acid-base catalysis. Other roles for histidines in substrate binding and/or in protein structure stabilization are possible as well. Two histidines are absolutely conserved in every ADP-glucose pyrophosphorylase sequenced to date. A third histidine is highly conserved in nine out of eleven sequences; the remaining two exist as glutamine. Site directed mutagenesis studies were carried out to replace each of these histidines with glutamine initially; additional residues asparagine, arginine and aspartate were substituted at the 83 position to further test its function.

EXPERIMENTAL PROCEDURES

Reagents. [^{32}P]pyrophosphate, [^{14}C]glucose-1-phosphate and [^{35}S]-dATP were purchased from DuPont-New England Nuclear. Enzymes for DNA manipulation and sequencing were from New England Biolabs or Boehringer Mannheim. Oligonucleotides were synthesized and purified by the Macromolecular Facility at Michigan State University. All other reagents were purchased as the highest quality available.

Bacterial strains and media. Bacterial strains used included *E. coli* MV1193 (Δ(lac-proAB) rpsL thi endA spcB15 hsdR4 Δ(srl-recA)306::Tn 10(tetr) F'[traD36 proAB+ lacIq lacZΔM15]), E. coli CJ236 (dut, ung, thi, rel A/pCj105 (Cmr)), and *E. coli* K12 G6MD3 (Hfr, his thi, Strr, Δ(mal - asd)).

[1] This research was supported in part by Department of Energy Research Grant DE-FGO2-93ER 20121.

[2] Present address: Department of Chemistry, Central Michigan University, Mt.Pleasant, MI 48859.

[3] Corresponding author. Fax: 517-353-9334. E-mail: preiss@pilot.msu.edu.

0006-291X/98 $25.00

His-83	5'-TG GTG CAG CAC ATT CAG CG-3'
	3'-AC CAC GTC GTG TAA GTG GC-5'
Gln-83	3'-AC CAC GTC GTT TAA GTG GC-5'
	*
Asn-83	3'-AC CAC GTC TTG TAA GTG GC-5'
	*
Arg-83	3'-AC CAC GTC GCG TAA GTG GC-5'
	*
Asp-83	3'-AC CAC GTC CTG TAA GTG GC-5'
	*
His-156	5'-TT ATC GAT CAC GTC GAA AA-3'
	3'-AA TAG CTA GTG CAG CTT TT-5'
Gln-156	3'-AA TAG CTA GTT CAG CTT TT-5'
	*
His-143	5'-CG GGC GAC CAT ATC TAC AA-3'
	3'-GC CCG CTG GTA TAG ATG TT-5'
Gln-143	3'-GC CCG CTG GTT TAG ATG TT-5'
	*

FIG. 1. Nucleotide sequences and encoded protein sequences of the ADP-glucose pyrophosphorylase gene in the regions of His-83, His-143, and His-156, and the synthetic oligonucleotides used to generate amino acid substitutions at these positions. The base substitutions are marked with asterisks.

G6MD3 cells were grown in enriched medium which contained 1.1% K_2HPO4, 0.85% KH_2PO4, 0.6% yeast extract, 0.2% glucose, pH 7.0. G6MD3 cell cultures also contained 50 mg/ml diaminopimelic acid, required for growth. MV1193 cells were grown in two media: M9 minimal medium containing 0.56% Na_2HPO4, 0.3% KH_2PO4, 0.05% NaCl, 0.1% NH_4Cl, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.001% thiamine, and 0.2% glucose, and LB medium containing 1% tryptone, 0.5% yeast extract, 0.5% NaCl. CJ236 cells were grown in LB medium containing 30 mg/ml chloramphenicol.

Site-directed mutagenesis. The gene for the native *E. coli* K12 ADP-glucose pyrophosphorylase was subcloned from the pOP12 plasmid (2) into M13mp18RF. The fragment used, the 1.9 kilobase *Hinc* II fragment of pOP12, contains the complete coding region for ADP-glucose pyrophosphorylase as well as the 474 base upstream region which encodes promoter site(s) necessary for expression of the gene(3). Site-directed mutagenesis was performed using the method of Kunkel (4,5). The mutant oligonucleotides used are shown in Fig. 1. All mutant DNAs were completely sequenced.

TABLE I

Specific Catalytic Activities in Crude Cell Extracts

	Specific activity (units/mg protein)
wild-type	5.40 ± 0.10
H156Q	1.43 ± 0.06
H143Q	2.69 ± 0.04
H83Q	3.70 ± 0.10
H83N	2.15 ± 0.05
H83R	0.15 ± 0.01
H83D	0.026 ± 0.002

Note. Reactions were performed at 37 °C as described under "Experimental Procedures". Data represent the averages of two identical experiments ± the average difference of the duplicates.

[a] One unit of enzyme activity is expressed as the amount of enzyme required to form one micromole of ATP per minute at 37 °C assayed in the pyrophosphorolysis direction as described under "Experimental Procedures."

Expression and purification of mutant and wild-type enzymes. For expression of wild-type and mutant enzymes, G6MD3 cells were infected with M13 phage carrying the desired gene.

Wild-type enzyme was purified according to established procedures (6, 7). Mutant enzymes, H83Q and H83N, were purified using these same procedures with the following modifications, found to prevent denaturation of these more labile proteins. The 60 °C heat step was replaced with a 35% ammonium sulfate saturation step in which enzyme remained in solution. This was followed by a 55% ammonium sulfate saturation step in which enzyme precipitated. Mutant enzyme was resuspended in 0.05 M glycylglycine, 1 mM EDTA, 5 mM DTT, pH 7.0 and dialyzed versus the same at 4 °C. This mutant enzyme sample was chromatographed on DEAE-Sepharose as de-

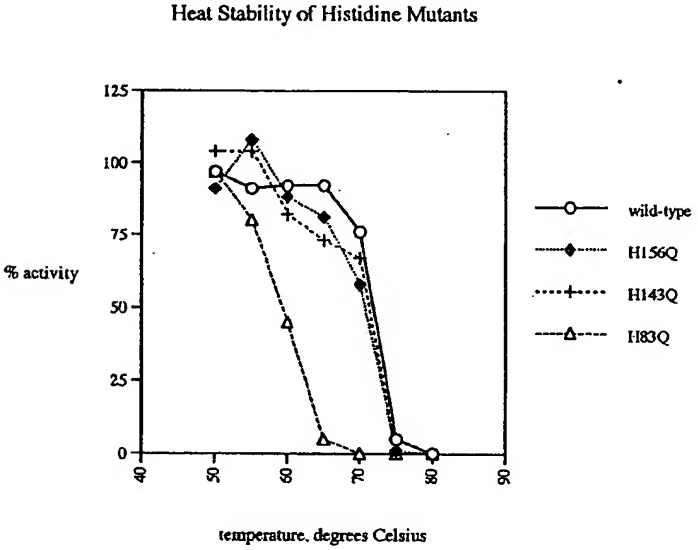


FIG. 2. Thermal stability comparison of wild-type, H156Q, H143Q, and H83Q proteins. Individual samples of G6MD3 cell extracts were heated simultaneously at the indicated temperatures for five minutes in the presence of 30 mM sodium phosphate, pH 7.0. After cooling on ice, samples were clarified by centrifugation and supernatants were assayed in the synthesis direction of assay. Symbols represent wild-type (O), H156Q (♦), H143Q (+), and H83Q (Δ).

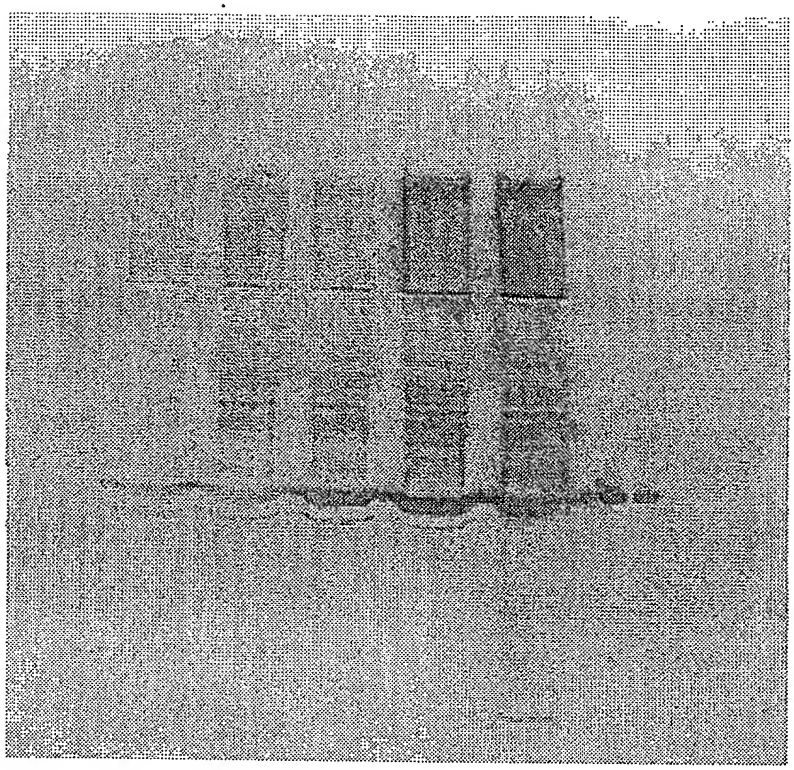


FIG. 3. Western Blot comparison of wild-type and mutant ADP-glucose pyrophosphoryase expression in G6MD3 cells. Extracts of expressed cells were subjected to SDS-PAGE; in each case 8 mg of protein was applied to the gel. Proteins were transferred to nitrocellulose and visualized using antiserum prepared against *E. coli* ADP-glucose pyrophosphorylase as described under "Experimental Procedures". Lane1 represents blank G6MD3 cells in which no ADP-glucose pyrophosphoryase was expressed. Lanes 2 - 6 represent expressed H83D, H83R, H83N, H83Q, and wild-type enzymes, respectively. Lane 7 represents purified wild-type ADP-glucose pyrophosphorylase.

scribed (6, 7). Following DEAE-Sepharose chromatography, the enzyme was concentrated and dialyzed as described, and then chromatographed directly on a Mono Q HR 5/5 Fast Protein Liquid Chromatography column. Purified mutant enzymes were desalted using an Econopak 10 desalting column (BioRad) equilibrated with 0.05 M Tris-HCl pH 7.5, 1mM EDTA, 0.5 mM DTT, 10 % glycerol. KCl was added immediately to a concentration of 1 mM and the enzyme quick-frozen in liquid nitrogen and stored at −80 °C. The Green A column was omitted during both mutant enzyme purifications. Each purified enzyme migrated as a single band (MW 50 kD) on SDS-PAGE.

Preparation of cell extracts for assay. Crude cell extracts of wild-type and mutant enzymes were prepared by resuspension of frozen, infected G6MD3 cells (0.5 grams) in 10 ml 0.05 M glycylglycine, 1 mM EDTA, 5 mM DTT, pH 7.0, on ice. Cells were disrupted by sonication using time and energy settings which were empirically found to be the minimum for complete cell disruption under these conditions. Cell extracts were clarified by centrifugation at 4 °C and assayed the same day of their preparation in all cases.

Enzyme kinetics. Enzymatic activity in the ADP-glucose synthesis direction at 37 °C was measured according to the method of Preiss, et al. (8). For assay of wild-type and mutant enzymes, reaction mixtures (final volume 200 ml) contained 0.1 mmol of [^{14}C] glucose-1-phosphate (specific activity 500 -1000 cpm/nmol), 0.3 mmol of ATP, 1.0 mmol of $MgCl_2$, 0.3 mmol of fructose-1,6-bisphosphate, 20 mmol of Hepes buffer, pH 7.0, and 100 mg of bovine serum albumin.

Enzymatic activity in the pyrophosphorylase direction at 37 °C was measured according to Preiss, et al. (8). Wild-type and mutant reaction mixtures (final volume 260 ml) contained 1.0 mmol of ADP-glucose, 0.43 mmol of [^{32}P]pyrophosphate (specific activity 4000 cpm/nmol), 2.0 mmol of $MgCl_2$, 0.78 mmol of fructose 1,6-bisphosphate, 10.0 mmol of Tris-HCl, pH 8.5, and 100 mg of bovine serum albumin.

Kinetic characterization. Kinetic data were plotted as intial velocity versus substrate or effector concentration. Data were replotted as double-reciprocal plots and the method of Wilkinson (9) was used to determine V_{max}. Sigmoidal plots were replotted as Hill plots to obtain kinetic constants. For sigmoidal data the following expressions for kinetic constants were used: $A_{0.5}$, $I_{0.5}$, and $S_{0.5}$, concentration of activator, inhibitor, or substrate, respectively, giving 50% maximal activation, inhibition, or maximal velocity. Duplicates were run in each case; kinetic constants are expressed as the mean ± the difference from duplicate determinations.

Thermal stability. Crude cell extracts were prepared as described above. Potassium phosphate, pH 7.0, was added to give a final concentration of 30 mM phosphate. Individual samples (50 μl volume) of crude extracts were heated for 5 min in a water bath equilibrated at the specified temperature, then immediately placed on ice. Samples were centrifuged, then assayed in the synthesis direction as described above.

Protein determination. Protein was assayed by the method of Smith, et al. (10) using bovine serum albumin as the standard.

SDS PAGE and Western blotting. SDS PAGE was performed in 10% gels using the method of Laemmli (11). Following electrophoresis, proteins were transferred to nitrocellulose membranes (12).

Staining was accomplished using rabbit antibody raised against the E coli ADP-glucose pyrophosphorylase, followed by alkaline phosphatase-labelled goat anti-rabbit antibody and subsequent visualization.

RESULTS

None of the three conserved histidines, H83, H143, or H156, are essential for catalytic activity in the *E. coli* enzyme. In each case when glutamine is substituted for histidine a functional enzyme results, though the activity is slightly lower than wild-type (Table I). An interesting, and quite critical observation of the H83Q mutant was that the original sonication conditions resulted in low specific activity measurements. These conditions used extended times of sonication (five 15 second pulses) at a high intensity level while keeping the sample on ice. It was found that a shorter period of sonication (two fifteen second pulses) at a low intensity setting (with sample on ice) resulted in crude extracts with wild-type levels of activity. Analysis of the heat stabilities of these mutants (Figure 2) illustrates the probable cause for the disparate H83Q specific activities; this mutant is heat labile in comparison to the wild-type enzyme or the other two histidine mutants. The initial low activity presumably resulted from denaturation during sonication.

Three additional substitutions at position 83 were prepared and analyzed: asparagine, arginine, and aspartate. Specific catalytic activities for these mutant enzymes in crude cell suspensions (prepared using minimum sonication conditions) indicate that asparagine is also a conservative change, but that arginine and aspartate are deleterious to enzyme function (Table I). Cell extracts used for catalytic assay were subjected to SDS PAGE and western blotting to assess whether the low specific activities were the result of low protein expression in the original host cultures (Figure 3). The western blot reveals that significant protein is expressed in each case, discounting lack of expression as a reason for the low specific activity. The western blot also reveals an antibody-stained protein band which migrates as a 45 kD fragment, in addition to the native enzyme band at 50 kD. This 45 kD band is present in all of the mutant lanes but is absent in the wild-type sample. This suggests that proteolytic processing is a factor in the mutant expressions.

Heat stability decreases as the substitution at position 83 progresses from histidine to glutamine to asparagine to arginine as shown in Figure 4. The aspartate mutant activity was too low to measure heat stability.

The two most heat stable mutants, H83Q and H83N, were purified, initially using procedures developed for purification of the wild-type enzyme. During the purifications difficulties were encountered in maintaining the stabilities of both these mutant enzymes. In particular, mutants were found to precipitate under conditions of low salt or low protein concentration, especially

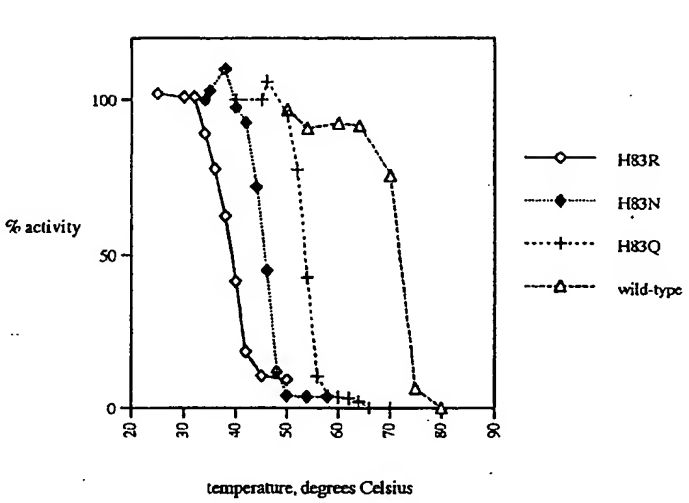


FIG. 4. Thermal stability comparison of wild-type, H83Q, H83N, and H83R proteins. Individual samples of G6MD3 cell extracts were heated at the indicated temperatures for five minutes in the presence of 30 mM sodium phosphate, pH 7.0. After cooling on ice, samples were clarified by centrifugation and supernatants were assayed in the synthesis direction of assay. Symbols represent wild-type (Δ), H83Q (+), H83N (♦), and H83R (◊).

during hours of dialysis following Fast Protein Liquid Chromatography on Mono Q. KCl at a final concentration of 1 mM was found to reverse or prevent this precipitation problem. H83Q and H83N purifications were repeated using a desalting column in place of dialysis at the final step of protein transfer into its storage buffer, to which was added 1 mM KCl. Despite these efforts to preserve mutant protein integrity, the specific catalytic activities of the two purified mutants were somewhat low in comparison to that of the wild-type enzyme: 18 and 28 units per mg for the asparagine and glutamine mutants, respectively, versus 74 units per mg for the wild-type enzyme (Table II).

It was also observed that the same purified enzyme sample (glutamine or asparagine mutant) showed a gradual loss of specific activity upon repeated freezings and thawings. Wild-type enzyme is resistant to this kind of activity loss.

Kinetic constants were measured for enzymes in the synthesis direction of assay, the metabolically significant direction. Kinetic constants (Table II) for substrates glucose 1-phosphate, ATP, and cofactor Mg^{2+} are similar in wild-type and mutant enzymes; a less than two-fold difference was obtained for each of these ligands. Similarly, the inhibitor AMP kinetic constant is unchanged as a result of histidine 83 substitution. Only the activator, fructose-1,6-bisphosphate, shows a large change in kinetic constant measurement ; $A_{0.5}$ values for both glutamine and asparagine mutants are on the order of 7-fold greater than the wild-type $A_{0.5}$.

TABLE II

Kinetic Constants of Wild-Type and Mutant ADP-Glucose Pyrophosphorylase

Glucose 1-phosphate	Wild-type	H83Q	H83N
Km (μM)	27.1 ± 0.9	22.8 ± 9.3	29.2 ± 4.3
ATP $S_{0.5}$ (μM)	334 ± 14	567 ± 5	535 ± 57
$MgCl_2$ $S_{0.5}$ (mM)	2.68 ± 0.09	4.56 ± 0.13	4.09 ± 0.08
AMP $I_{0.5}$ (μM)	89.5 ± 4.5	86.5 ± 14.5	90.8 ± 10.2
Fru 1,6-bis-P $A_{0.5}$ (μM)	33 ± 7	240 ± 32	178 ± 22
V_{max} (units/mg)	74.0 ± 9.0	28.4 ± 0.3	18.0 ± 1.1

Note. Reactions were performed at 37 °C in the synthesis direction of assay as described under "Experimental Procedures". Data represent the average of two identical experiments ± the average difference of the duplicates.

DISCUSSION

No absolutely conserved histidine is essential for catalysis in the *E. coli* ADP-glucose pyrophosphorylase enzyme. In fact, of the three conserved histidines, two may be replaced by glutamine with little change in specific catalytic activity (as measured in crude cell suspensions), and no change in heat stability. These residues, at positions 143 and 156 in the *E. coli* enzyme, must not function in the catalytic mechanism, nor be essential to protein structure stabilization.

Histidine 83 is also not essential for catalysis. Conservative replacements by glutamine or asparagine do not affect specific catalytic activity, thereby ruling out the possibility of His-83 involvement in catalysis. Furthermore, these substitutions have negligible effect on substrate, cofactor, and inhibitor binding, as judged by the similarities between wild-type and mutant kinetic constants measured for these ligands. However, perturbation of the activator (fructose 1,6-bis-phosphate) binding site constant does occur with these substitutions.

Stabilization of protein tertiary structure is concluded to be the reason for histidine conservation at the 83 position. The decrease in thermal stability and the increase in proteolytic susceptibility for the mutants provides support for this conclusion. Several studies suggest that thermal stability is a key determinant of proteolytic susceptiblity in the cell (13). Certainly, proteases show a preference for denatured proteins *in vitro* (14).

In light of this, it is not surprising that the "conservative" mutants H83Q and H83N were difficult to isolate in stable form and that precipitation and loss of activity ocurred during the purification process. The specific catalytic activities measured for these purified forms can be considered to be minimum values until purification conditions are found during which mutant stabilities can be maintained.

It is not possible to conclude whether His-83 is directly involved in fructose 1,6-bis-phosphate (FBP) binding. Previous covalent modification and site-directed mutagenesis studies identified residues at the N-terminus of the bacterial enzyme which are involved in FBP binding (15,16,17). Since His-83 is also located in this region it would not be surprising if it were to be part of the FBP site. However, perturbation of the FBP site in His-83 mutants may simply be a secondary effect from localized destabilization of tertiary structure in the FBP binding domain by a distant substituted residue 83. This would be reasonable given the effect of His-83 mutations on overall protein stability. It is also difficult to rationalize why an amino acid involved in bacterial enzyme FBP binding would be so well conserved in organisms which have evolved a different activator binding specificity. His-83 may therefore be conserved solely to preserve the folded, active state of the protein.

Efforts are continuing to probe the functional roles of other absolutely conserved amino acids in this enzyme. Recent success in the growth of crystals of E. coli enzyme suitable for x-ray diffraction will make it possible to evaluate whether His-83 does or does not participate directly in FBP binding.

REFERENCES

1. Smith-White, B. J., and Preiss, J. (1992) *J. Mol. Evol.* **34**, 449–464.
2. Okita, T. W., Rodriguez, R. L., and Preiss, J. (1981) *J. Biol. Chem.* **256**, 6944–6952.
3. Romeo, T., and Preiss, J. (1989) *J. Bacteriol.* **171**, 2773–2782.
4. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 488–492.
5. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382.
6. Ozaki, H., and Preiss, J. (1972) *Methods Enzymol.* **28**, 409–413.
7. Kumar, A., Ghosh, P., Lee, Y. M., Hill, M. A., and Preiss, J. (1989) *J. Biol. Chem.* **264**, 10464–10471.
8. Preiss, J., Shen, L., Greenberg, E., and Gentner, N. (1966) *J. Biol. Chem.* **5**, 1833–1845.
9. Wilkinson, G. N. (1961) *Biochem. J.* **80**, 324–332.
10. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Kenk, D. C. (1985) *Anal. Biochem.* **150**, 76–85.
11. Laemmli, U. K. (1970) *Nature* **227**, 680–685.
12. Burnette, W. W. (1981) *Anal. Biochem.* **112**,, 195–203.
13. Parsell, D. A., and Sauer, R. T. (1989) *J. Biol. Chem.* **264**, 7590–7595.
14. Pace, C. N., and Barrett, A. J. (1984) *Biochem. J.* **219**, 411–417.
15. Parsons, T., and Preiss, J. (1978) *J. Biol. Chem.* **253**, 6197–6202.
16. Carlson, C., and Preiss, J. (1981) *Biochemistry* **20**, 7519–7528.
17. Gardiol, A, and Preiss, J. (1990) *Arch. Biochem. Biophys.* **280**, 175–180.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 1999, p. 4369–4374
0099-2240/99/$04.00+0

Vol. 65, No. 10

Identification of *Bacillus thuringiensis* Delta-Endotoxin Cry1C Domain III Amino Acid Residues Involved in Insect Specificity

RUUD A. DE MAAGD,[1]* PETRA BAKKER,[1] NIKOLA STAYKOV,[2] STEFAN DUKIANDJIEV,[2] WILLEM STIEKEMA,[1] AND DIRK BOSCH[1]

Department of Molecular Biology, DLO-Centre for Plant Breeding and Reproduction Research (CPRO-DLO), 6700 AA Wageningen, The Netherlands,[1] and Department of Plant Physiology and Molecular Biology, University of Plovdiv, Plovdiv, Bulgaria[2]

Received 20 May 1999/Accepted 13 July 1999

Cry1C domain III amino acid residues involved in specificity for beet armyworm (*Spodoptera exigua*) were identified. For this purpose, intradomain III hybrids between Cry1E (nontoxic) and Cry1E-Cry1C hybrid G27 (toxic) were made. Crossover points of these hybrids defined six sequence blocks containing between 1 and 19 of the amino acid differences between Cry1E and G27. Blocks B, C, D, and E of G27 were shown to be required for optimal activity against *S. exigua*. Block E was also required for optimal activity against the tobacco hornworm (*Manduca sexta*), whereas block D had a negative effect on toxicity for this insect. The mutagenesis of individual amino acids in block B identified Trp-476 as the only amino acid in this block essential, although not sufficient by itself, for full *S. exigua* activity. In block D, we identified a seven-amino-acid insertion in G27 that was not in Cry1E. The deletion of either one of two groups of four consecutive amino acids in this insertion completely abolished activity against *S. exigua* but resulted in higher activity against *M. sexta*. Alanine substitutions of the first group had little effect on toxicity, whereas alanine substitutions of the second group had the same effect as its deletion. These results identify groups of amino acids as well as some individual residues in Cry1C domain III, which are strongly involved in *S. exigua*-specific activity as well as sometimes involved in *M. sexta*-specific activity.

Bacillus thuringiensis is a gram-positive bacterium which, during sporulation, produces parasporal crystalline inclusions consisting of one or several delta-endotoxins, or Cry proteins, which have insecticidal properties. The *cry* gene family is a large, still-growing family of homologous genes, and each gene encodes a protein with activity against only one or a few insect species. To be toxic, *B. thuringiensis* crystals have to be ingested by the insect. In the insect midgut, the crystals are dissolved, releasing their constituent Cry proteins as protoxins with molecular masses of 70 to 130 kDa. Midgut proteases subsequently trim the protoxins, resulting in a truncated, N-terminal fragment with a mass of approximately 65 kDa, which is the activated toxin. The activated toxin then binds to specific receptors on the surface of the midgut epithelial cells and penetrates the cell membranes, forming pores and killing the epithelial cells by colloid osmotic lysis (reviewed in reference 13). Much, although not all, of the observed specificity of the Cry proteins is determined by the interaction of the toxin with specific receptors, if present, on the midgut epithelial cell membranes (14).

Progress has been made both in determining the three-dimensional structure of the toxin molecule and in identifying the primary sequences involved in specificity and receptor binding, allowing the study of structure-function relationships. The published structures of two delta-endotoxins show a three-domain structure (8, 10). The N-terminal domain I consists of seven alpha-helices and is thought to be responsible for insertions into the insect cell membrane and to be involved in pore formation. The more variable domain II contains the primary sequences that have been shown to be involved in insect specificity and in high-affinity binding (13). Domain II is therefore assumed to be involved in actual interactions with receptors and, to a large extent, in determining specificity through that process. The function of the C-terminal domain III at the molecular level was unknown until recently, although a variety of mutagenesis and recombination experiments have shown that it can be involved in specificity. Most notably, an exchange of domain III (by in vivo recombination) between toxins with differing specificities can change the specificity and even create new hybrid toxins with increased activity (2, 5, 7, 11). Our group, as well as others, has shown that a substitution of domain III alters the binding of toxins to putative receptors on ligand blots of brush border membrane vesicle proteins separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1, 5, 6, 9). Recently, using various different binding assays, we have shown that domain III of Cry1Ac is involved in *N*-acetylgalactosamine (GalNAc)-inhibited binding to the putative receptor in *Manduca sexta*, aminopeptidase N (APN) (4). These recent studies make it increasingly likely that the role of domain III in determining insect specificity is probably correlated with its ability to bind, with or without domain II, to specific receptors on the surfaces of the gut epithelial cells of the target insect.

In contrast to the numerous studies on the role of domain II sequences in specificity and target membrane binding, few data on the role of domain III amino acids in these functions are currently available. A mutation of three amino acids in domain III of Cry1Ac was shown to strongly decrease GalNAc-inhibited binding by Cry1Ac to *M. sexta* membranes, implicating a role of these residues in a physical interaction between domain

* Corresponding author. Mailing address: Department of Molecular Biology, DLO-Centre for Plant Breeding and Reproduction Research (CPRO-DLO), P.O. Box 16, 6700 AA Wageningen, The Netherlands. Phone: 31 317 477128. Fax: 31 317 418094. E-mail: R.A.deMaagd@cpro.dlo.nl.

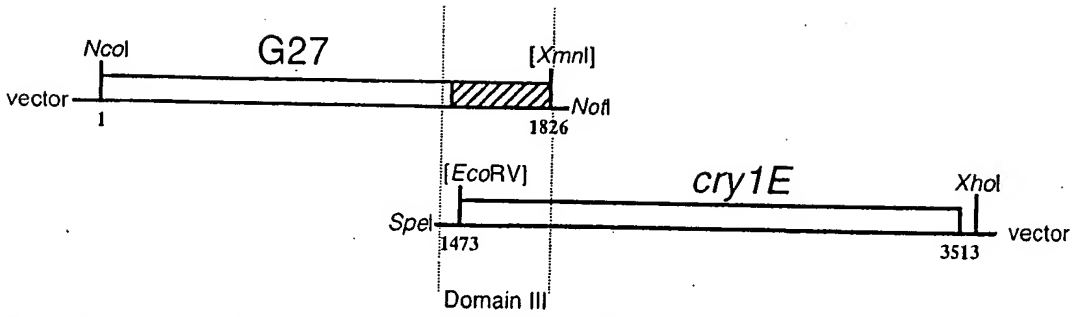


FIG. 1. Schematic representation of the G27 gene-*cry1E* tandem plasmid pHK4. Open and dashed boxes represent Cry1E- and Cry1C-encoding sequences, respectively. Locations of restriction enzyme sites used for cloning are indicated. Dotted vertical lines indicate the borders of the domain III-encoding sequences. For clarity, the overlapping regions of G27 and Cry1E are aligned vertically, and the polylinker between the two genes is shown as cut by *Not*I and *Spe*I.

III and GalNAc or a similar sugar on APN (3). In yet another study, a mutation of two amino acids in domain III of Cry1Ac was shown to diminish its toxicity for *M. sexta* as well as for *Heliothis virescens*, while binding to putative receptors on ligand blots of brush border membrane proteins from those insects had decreased as well. In the latter study, the effects on toxicity for *M. sexta* were larger than those on toxicity for *H. virescens* (1).

Exchanging domains III of Cry1C, which is active against *Spodoptera exigua*, and Cry1E, which is not active, resulted in the *S. exigua*-active Cry1E-Cry1C hybrid G27 (2). These results indicated that domain III of Cry1C is involved in specificity for *S. exigua*. This was confirmed by our finding that domain III of Cry1C can also render the inactive toxin Cry1Ab active against *S. exigua* (5). In this study, we characterized intradomain hybrids between domains III of Cry1E and Cry1E-Cry1C hybrid protein G27 (homologue scanning) in order to more precisely determine which parts of domain III of Cry1C are involved in *S. exigua*-specific activity. In addition, we identified individual amino acids that are involved in this specificity by mutagenesis.

MATERIALS AND METHODS

Plasmids. Cry1E expression plasmid pBD160, expression plasmid pBD151 containing *Xmn*I (base 1853)-truncated *cry1C*, and *cry1E-cry1C* hybrids G27, H7, H8, and H17 were described earlier (2). For the recombination of the 3′ end of the hybrid G27 gene with *cry1E*, tandem plasmid pHK4 was constructed (Fig. 1). For this purpose, the 3′ part of the G27 gene, a *Bst*BI-*Xho*I fragment (from base 1488 of *cry1C* to its end) was replaced by the corresponding fragment of pBD151, which contains the truncated 3′ end of the active-Cry1C-encoding DNA (up to base 1835) followed by the polylinker of pBluescript SK(+). Subsequently, an *Eco*RV-*Xho*I fragment of pBD160 containing the 3′ end of the Cry1E toxin-encoding DNA from base 1473 to the end of the gene was cloned into the vector fragment of *Sma*I-*Xho*I-digested pBD151.

Mutagenesis. All mutations were made with the QuickChange kit (Stratagene) and complementary mutagenic oligonucleotides. Mutant plasmids pNS1, pNS20, pNS22, and pB16 were derived from *cry1E*. pNS2, pNS3, pNS21, and pNS23 were derived from H8. All mutations were checked by DNA sequencing.

In vivo recombination. *E. coli* JM101 (*recA*$^+$) was transformed with pHK4, and then plasmid DNA was isolated. To select for recombinant plasmids, DNA was digested with *Not*I and *Spe*I, which have unique sites in the polylinker region between the G27- and Cry1E-encoding parts (Fig. 1). Digested DNA was transferred to *E. coli* XL-1 by transformation, and transformants were screened for recombination events by restriction analysis of isolated DNA.

Toxin production, purification, and bioassays. All protoxins were produced in *E. coli* XL-1, and trypsin-activated toxins were purified by fast protein liquid chromatography as described earlier (2). The toxicities of the proteins were tested by spreading toxin dilutions on artificial diet. Neonate larvae of *S. exigua* were used, and their mortality was scored after 6 days at 28°C. For *M. sexta* bioassays, 1-day-old larvae were used and their mortality was scored after 6 days at 28°C. The concentrations with 50% lethality (LC_{50}s) and 95% fiducial limits were estimated by a Probit analysis of results from three or more independent experiments, using the PoloPC computer program (12).

RESULTS

Substitution of N-terminal domain III sequences. Our original Cry1E-Cry1C hybrid protein G27 consisted of domains I and II, as well as the most N-terminal part of domain III of Cry1E, while the rest of the protein, comprising the major C-terminal part of domain III as well as the protoxin-specific fragment, was derived from Cry1C (Fig. 2 and 3). In contrast to Cry1E, Cry1C and G27 are toxic to *S. exigua*, indicating that the Cry1C-derived part of G27 contains sequences that are specifically required for activity against this insect. Besides G27, the original set of Cry1E-Cry1C hybrids contained a number of hybrids with smaller contributions of domain III from Cry1C (Fig. 2 and 3, hybrids H8, H17, and H7) (2). The locations of the crossover points define blocks (A through F) of the domain III sequence which were replaced in the various hybrids produced in that study (Fig. 2 and 3). Whereas G27 has high activity against *S. exigua*, the activities of the other hybrids decreased rapidly as the crossover point moved further into domain III (i.e., less of domain III from Cry1C) (Fig. 3). A replacement in G27 of only three amino acids (block B) by their corresponding amino acids from Cry1E (H8) decreased activity approximately 10-fold. A further replacement of two amino acids in block C (H17) decreased activity again fivefold. Whereas both H8 and H17 were still distinctively more toxic than Cry1E, a further replacement of block D (H7) left no detectable activity against *S. exigua*. At the same time, G27, H8, and H17 were equally active against *M. sexta* (Fig. 3), suggesting that blocks B and C are involved in specific activity against *S. exigua*. H7 also had no activity against *M. sexta*, suggesting that a simultaneous substitution of blocks B to D destroys biological activity in a less specific way.

Substitution of C-terminal domain III sequences. To test the contributions of C-terminal amino acids in domain III and in the protoxin of Cry1C to the activity of G27, a new tandem plasmid for in vivo recombination, pHK4, was constructed. This plasmid contains the gene encoding G27, which was truncated at the end of the domain III-encoding part, followed by a polylinker and the parts of the *cry1E* gene encoding domain III and its protoxin tail to the 3′ end (Fig. 1). This allowed in vivo recombination between the domain III-encoding parts of the G27 gene and *cry1E*, in effect replacing 3′ sequences of the G27 gene with those of *cry1E*. New recombinant toxins were screened at the DNA level by restriction analysis and sequencing to obtain crossover events throughout domain III without prior selection for soluble-protoxin-encoding genes of hybrid toxins. Hybrids BS21, NS6, NS8, and BS22 had increasing numbers of their C-terminal amino acids replaced by the cor-

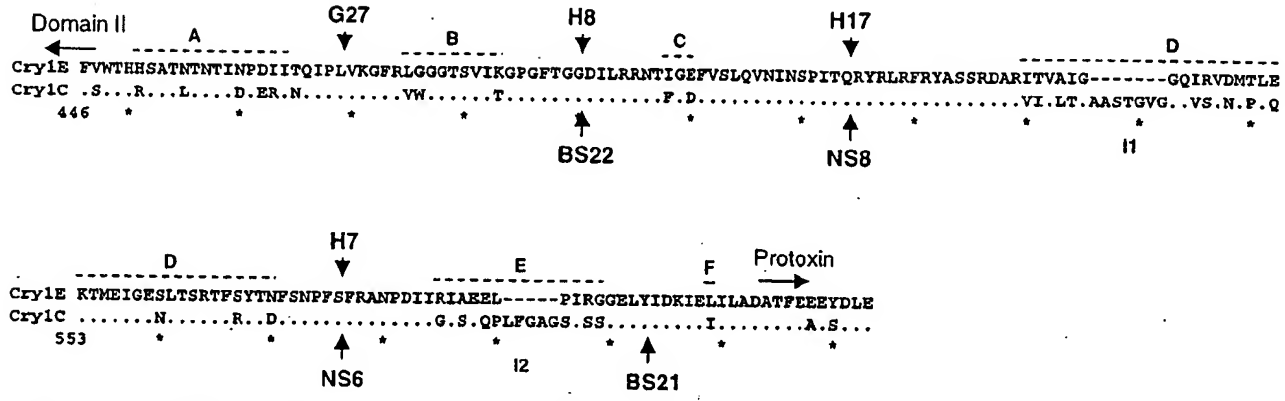


FIG. 2. Amino acid alignment of domains III of Cry1E and Cry1C. Identical amino acids of Cry1C are depicted by dots. Arrowheads indicate approximate locations of crossover sites in Cry1E-Cry1C hybrids (above) and in G27-Cry1E-hybrids (below). Because of the homology in these regions, exact locations of the crossover sites could not be determined. On top of the alignment, the relative positions of amino acid blocks A through F are depicted by dashes. I1 and I2 indicate the positions of two insertions in Cry1C. Amino acid numbering follows that of G27.

responding Cry1E residues (Fig. 3). Only hybrids BS21 and NS6 produced soluble protoxins, and they also produced, upon trypsin treatment, a stable activated toxin. Hybrid BS21 had a crossover site near the C-terminal end of domain III, making it a protoxin that consisted mainly of Cry1E sequences with the exception of a part of domain III which was derived from Cry1C (Fig. 3). Trypsin-activated BS21 protein retained the activity of G27 against *S. exigua*, indicating that specificity for *S. exigua* is conferred by just this part (comprising blocks B, C, D, and E) of domain III of Cry1C (Fig. 3). A further replacement of domain III C-terminal amino acids (block E), as seen with hybrid NS6, decreased the activity of the resulting toxin against both *S. exigua* and *M. sexta* (Fig. 3). Still further replacement by Cry1E sequences (blocks D and C, respectively), as seen with hybrids NS8 and BS22, resulted in hybrid genes for which no soluble protoxin could be recovered. An analysis of whole-cell lysates showed that both genes did express proteins of the correct, expected size, although in amounts smaller (approximately 10%) than that of G27 (results not shown). Nonetheless, no protein was recovered during attempts to sol-

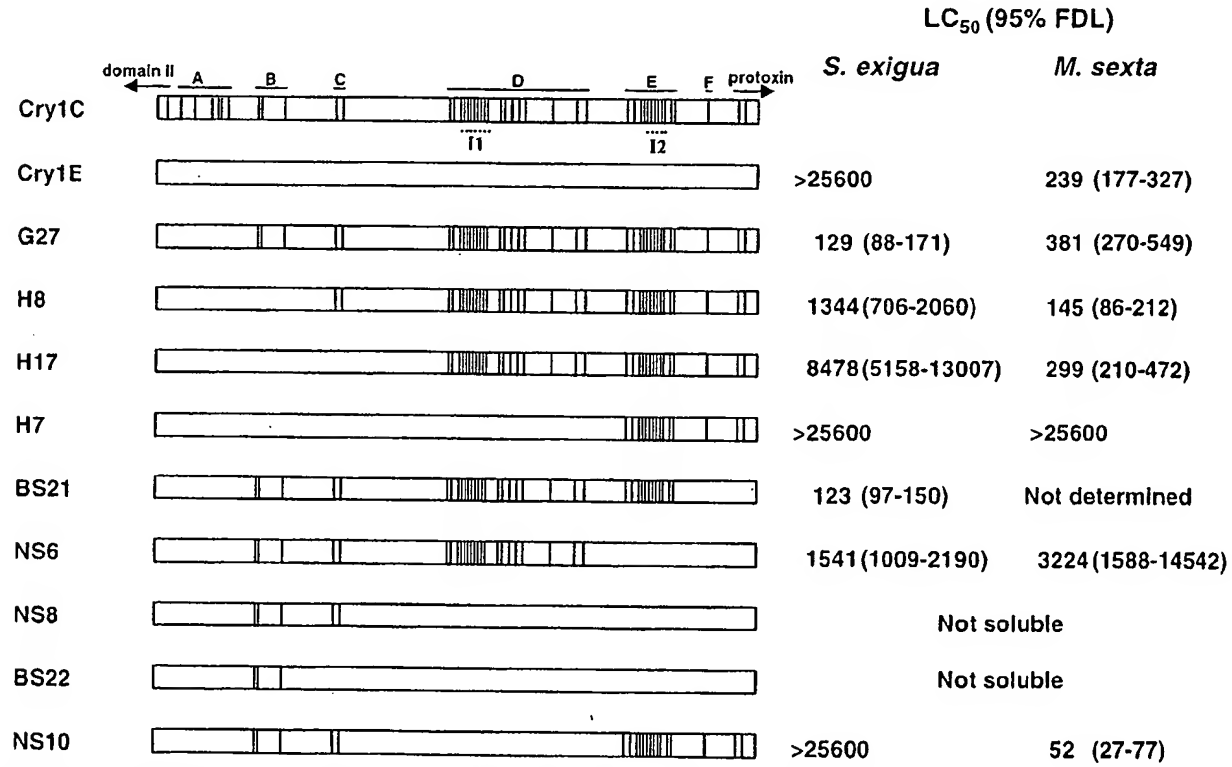


FIG. 3. Schematic representation of domains III of wild-type and hybrid toxins used in this study. Vertical lines represent the amino acid residues derived from Cry1C, which are different in Cry1E. At the top, the relative positions of amino acid blocks A through F are depicted by horizontal lines. I1 and I2 indicate the positions of two insertions. Toxicity for *S. exigua* and *M. sexta* is represented as LC_{50}, in nanograms per square centimeter. FDL, fiducial limits.

TABLE 1. Amino acid sequences of domain III block B mutants and their toxicities for *S. exigua*

Toxin	Sequence	Background[a]	LC_{50} (95% FDL)[b]
Cry1E	^{473}FRLGGGTSVIKGPG^{486}	Cry1E	>25,600
G27	..VW.....T...	H8	129 (88–172)
H8	.....	H8	1,344 (706–2,060)
BS22	..VW.....T...	Cry1E	ND
NS1	..VW.....	Cry1E	ND
NS22	..V.....	Cry1E	>25,600
NS20	...W.....	Cry1E	14,000
PB16	.....T...	Cry1E	>25,600
NS2	..VW.....	H8	103 (76–131)
NS23	..V.....	H8	>12,800
NS21	...W.....	H8	80 (56–103)
NS3	.....T...	H8	2,042 (888–3,284)

[a] Toxin from which domain III (excluding block B) was derived.
[b] Toxicity is indicated as LC_{50}, in nanograms per square centimeter. FDL, fiducial limits. ND, not determined.
[c] Reliable fiducial limits could not be determined.

ubilize these protoxins. Therefore, they were not tested further for toxicity.

Mutagenesis of block B amino acids. A comparison of the toxicities of G27 and H8 showed that a replacement of only three amino acids (block B) by their corresponding Cry1E residues severely and specifically affected activity against *S. exigua*, but not activity against *M. sexta* (see above). To further study the contribution of each of these three amino acids to its specificity for *S. exigua*, we made a number of mutants of Cry1E in which one or two of these amino acids were replaced by their equivalents from Cry1C (Table 1). Our results with hybrid BS22 (above) had already shown that a simultaneous replacement of all three would result in the inability to recover a soluble protoxin. Replacements of neighboring Leu-475 and Gly-476 (mutant NS1) also resulted in no soluble protoxins. Individual replacements Leu475Val, Gly476Trp, and Lys483Thr to produce mutants NS22, NS20, and PB16, respectively, did result in the recovery of soluble protoxins. Only one mutant, NS20, was significantly more active against *S. exigua* than was Cry1E (Table 1), although only at a low level. This indicated that although one or more of these Cry1C amino acids was essential for specific activity, no single one was sufficient to raise the activity to a level comparable to that of G27. As the presence of other specific Cry1C amino acids was apparently required, we repeated the mutagenesis described above in the background of hybrid H8, which contained the rest of domain III from Cry1C. The simultaneous replacements of Leu475Val and Gly476Trp, as well as the replacement of only one of each of the three amino acids (NS2, NS23, NS21, and NS3), resulted in the production of soluble protoxins. However, only NS2 and NS21 (replacements Leu475Val and Gly476Trp, and Gly476Trp alone, respectively) were as toxic to *S. exigua* as was G27, while NS23 and NS3 were only as toxic as H8 at best (Table 1). This shows that of the three amino acids in block B, only Trp-476 is required, and sufficient, for optimal activity against *S. exigua*.

Mutagenesis of block D amino acids. The initial experiments, which compared hybrids H17 and H7 (above), indicated that block D, which is of a different origin in these hybrids, plays a role in activity against *S. exigua*. However, because H17 already has low (but significant) activity, and because a simultaneous substitution of blocks B to D (H7) apparently destroys biological activity, the exact level of contribution of block D to toxicity for *S. exigua* is unclear. Interestingly, a major part of the amino acid differences between Cry1C and Cry1E domains III is concentrated in this block. To study the role of this block more closely, we produced a mosaic of hybrids NS8 and H7, called NS10, which effectively resulted in a replacement of block D amino acids in G27 by their equivalents from Cry1E (Fig. 3). In contrast to hybrid NS8, which yielded no soluble protoxin, hybrid NS10 yielded a soluble protoxin and a stable activated toxin upon trypsin treatment. This toxin had no detectable activity against *S. exigua*, but it was active against *M. sexta* (Fig. 3), indicating a major contribution of this block to *S. exigua*-specific activity. Curiously, NS10 was approximately seven times more toxic to *M. sexta* than was G27.

An alignment of the Cry1C and Cry1E domain III amino acid sequences (Fig. 2) shows two insertions of seven and five amino acids, respectively, in Cry1C. The first insertion is in block D, around residue 540, and was therefore studied more closely by means of mutagenesis experiments. Mutagenesis was performed to result either in deletions or in alanine substitutions of two blocks of four amino acids in this insertion, as shown in Table 2. All four mutations, both alanine substitutions as well as deletions, resulted in the production of soluble protoxins and stable activated toxins upon trypsin treatment, and these were tested for insecticidal activity. Both deletions (NS13 and NS16) resulted in the loss of activity against *S. exigua*. Alanine substitution of the second group (NS15) also lead to loss of activity, but substitution of the first group (NS14) only slightly affected activity against *S. exigua*. All four mutants were toxic to *M. sexta* and in all cases were three to four times more active than G27 (Table 2). These results show that at least part of the first insertion region in domain III of G27 is essential for activity against *S. exigua*, but that it simultaneously has a negative effect on activity against *M. sexta*.

DISCUSSION

Cry1C is the most toxic, natural *B. thuringiensis* delta-endotoxin for *S. exigua* described so far. Domain III of Cry1C is an important factor in this toxicity, as a transfer of Cry1C domain III can change the inactive toxins Cry1E and Cry1Ab into *S. exigua*-active hybrid toxins (2, 5). In this study, we made use of the relatively high homology between the domain III-encoding parts of *cry1C* and *cry1E* to create a variety of intradomain hybrids by in vivo recombination in order to more precisely identify the parts of the Cry1C domain III which are responsible for specificity for *S. exigua*. Study of these intradomain hybrids not only allows conclusions on the requirements for activity against *S. exigua* but also provides information about interactions between different parts of the primary sequence of domain III.

For the convenience of the discussion we divided the amino acid differences between the domains III of Cry1C and of Cry1E into six blocks, each containing between 1 and 19 residue differences, which are defined by the crossover sites of the hybrid toxins used in this study (Fig. 2 and 3). A comparison of

TABLE 2. Amino acid sequences of domain III block D mutants of G27 and their toxicities to *S. exigua* and *M. sexta*

Toxin	Sequence[a]	LC_{50} (95% FDL)[b] for:	
		S. exigua	*M. sexta*
G27	^{530}VIVLTGAASTGVGGQT^{545}	129 (88–172)	381 (270–549)
NS13	....---------.....	>20,000	56 (29–87)
NS16	.....---------....	>20,000	100 (83–119)
NS14	....AA.....	244 (170–338)	72 (52–98)
NS15	.....AAAA....	>25,600	115 (89–148)

[a] Amino acid deletions are indicated by dashes.
[b] Toxicity is indicated as LC_{50}, in nanograms per square centimeter. FDL, fiducial limits.

```
                    β15        β16
Cry1Aa  495GGDILRRTSPGQISTLRVNITAPL518
Cry1Ac  495...LV.LN.S.NNIQN.GY.EV.I518
G27     489....L.RNTF.DFVS.Q...NS.I512
```

FIG. 4. Alignment of domain III amino acid sequences encompassing block C of G27. Cry1Aa amino acids in β-sheets β15 and β16 are in italics. Amino acids identical to those of Cry1Aa are depicted by dots. Cry1Ac amino acids implicated in toxicity or specificity and G27 amino acids of block C (see Discussion) are bold and underlined.

Cry1C, Cry1E, and BS21 showed that only blocks B through E of Cry1C, or parts thereof, are essential for a high level of activity against *S. exigua* (Fig. 3). As we have found no soluble, trypsin-stable hybrid toxin that contains all of domain III of Cry1C, including block A, we can draw no conclusions on the role of this block. The replacement of the C-terminal (protoxin-specific) part of G27 by a Cry1E sequence (hybrid BS21) had no effect on toxicity. This shows that the difference between G27 and Cry1E toxicities is not likely to be caused by differences in processing by trypsin, which could result from differences in the protoxin-encoding parts of these genes. Furthermore, the conservative substitution of block F (Ile609Leu) had no effect on toxicity.

Replacements of block B were shown to have a strong negative effect on the activity of G27 against *S. exigua* (Fig. 3, H8 and G27). Further study of the role of the three individual residue differences in this block revealed that the replacement of only a single amino acid, Trp-476, could account entirely for this effect (Table 1). However, this residue could exert maximal positive effect on the activity against *S. exigua* only in conjunction with other Cry1C sequences from blocks C through E (Table 1, NS20 and NS21). Interestingly, the simultaneous placement of the neighboring Val-475 and Trp-476 into a Cry1E background, as well as of the whole block B (BS22) or blocks B and C (NS8), yielded no soluble protoxin. This suggests that these two-amino-acid substitutions may interfere with proper folding in the absence of other complementary Cry1C domain III sequences. This may have led to degradation in *E. coli* by intracellular proteases, resulting in the inability to recover any soluble protoxin. The presence of either Cry1C block D or block E apparently restored proper folding, as hybrids NS6 and NS10 are soluble and stable upon trypsin treatment.

Either one or both of the amino acids (Phe-498 and Glu-500) of block C are involved in activity against *S. exigua*, as their replacement further decreased toxicity (Fig. 3, H17). An alignment of domains III of Cry1Ca, Cry1Ac, and Cry1Aa (for which the three-dimensional structure has been determined) (8) (Fig. 4) placed these two amino acids in approximately the same location as two Cry1Ac amino acids (serines 503 and 504) that were shown to play a role in toxicity for *M. sexta* and *H. virescens* (1) and close to the location of the three Cry1Ac amino acids involved in GalNAc-inhibited binding to *M. sexta* APN (3).

A replacement of block D in G27 (Fig. 3, NS10) resulted in the complete destruction of activity against *S. exigua* without negatively affecting activity against *M. sexta*, indicating a major role of this block, which contains the most residue differences, in specificity. These differences are concentrated in one area around a seven-amino-acid insertion in Cry1C (around residue 540), on which we focused in further experiments. As homology between Cry1C and Cry1E in this area is low, the validity of the alignment around that area, and hence the exact location of the insertion (if continuous) in the sequence, as shown in Fig. 2, is uncertain. Therefore, we chose to study, by deletion and alanine substitution, two consecutive stretches of four amino acids (Table 2). Both deletions completely destroyed toxicity for *S. exigua*, indicating that either these residues are directly involved in an interaction with the target insect or the deletion of these residues perturbs the tertiary structure of the region, which could be essential for that interaction. Alanine substitutions of the first two amino acids (Tyr534Ala and Gly535Ala) of the first four-amino-acid stretch had only a slight effect on toxicity. On the other hand, alanine substitutions of the second four-amino-acid stretch (Ser-538 to Val-541) also completely destroyed activity against *S. exigua*, indicating that the identity of one or more of these four residues is a major determinant of specificity for *S. exigua*.

Initially, the toxicities of the hybrids and mutants to *M. sexta* were considered controls for checking to what extent changes affected *S. exigua*-specific activity as opposed to more general, nonspecific changes, which could result from suboptimal protein function due to, for example, improper folding. Since Cry1C and Cry1E are approximately equally active against *M. sexta*, any functional hybrid toxin derived from a combination of these two would be expected to be as active as G27 against *M. sexta*, or less so in the case of incompatibility of the different parts leading to misfolding. Surprisingly, hybrid NS10, in which block D of G27 was replaced by that of Cry1E, was significantly more active against *M. sexta* (approximately sevenfold) than G27 (Fig. 3). Also, the mutations within this block, described in Table 2, all led to higher toxicity for *M. sexta*. This indicates not only that block D of G27, or a part thereof, is strongly involved in high activity against *S. exigua* but also that its replacement in G27 by the corresponding block from Cry1E increases the toxicity of the resulting hybrid for *M. sexta*. A similar effect is achieved by short deletions or, to a lesser extent, by alanine substitutions within that block.

A replacement of block E of G27 by the corresponding block of Cry1E (hybrid NS6) severely diminished activity against both *S. exigua* and *M. sexta*. At present, we have no explanation for this effect. Although the protoxin of NS6 was readily solubilized and appeared to yield a stable toxin upon trypsin treatment, the combination of Cry1C and Cry1E parts may form a product which is less stable in the gut of both insects, leading to lower toxicity regardless of the insect species. This makes it impossible to determine the specificity-determining function of this block as a whole, although it may still be possible in the future to identify individual amino acid substitutions that affect activity against *S. exigua* and not that against *M. sexta*.

In conclusion, we have identified Cry1C domain III amino acids in larger blocks (blocks D and E) as well as in small blocks or groups (block C and the insertion in block D) and a single amino acid (Trp-476), which have a strong positive effect on activity against *S. exigua*. Some blocks and residue groups simultaneously have a negative effect on activity against *M. sexta* (block D), whereas block E in G27 is required for high activity against both insects. One might speculate that at least some of the identified specificity-determining residues are involved in binding, which has been shown to be a function of another domain III, i.e., that of Cry1Ac (3, 4). It is interesting to speculate about the location of the *S. exigua*-specific residues in the three-dimensional structure of domain III of Cry1C. Although no X-ray-diffraction-derived structure is available for this toxin, sequence alignment with Cry1Aa allows the prediction of its structure, to some extent. This places Trp-476 at the C-terminal end of beta-strand 13b and, in close proximity, places Phe-498 and Glu-500 between beta strands 15 and 16. The Cry1C-specific insertion in block D which is involved in specificity for *S. exigua* may lead to an increase in the length of the loop between beta strands 18 and

19, although structural changes in this area may be more extensive. Interestingly, all of these *S. exigua* specificity-determining residues appear to be located on the slightly concave outer beta sheet of domain III (8). Residues in the same sheet were implicated in the GalNAc-binding capacity of domain III of Cry1Ac (3). Our preliminary results have shown that domain III of Cry1C in G27 does increase the affinity of binding to *S. exigua* membranes, compared with Cry1E (results not shown). Future detailed studies of binding properties of hybrids and mutants may further clarify the functions of the identified residues at the molecular level.

ACKNOWLEDGMENTS

We are grateful to Ine Derksen, (Dept. of Virology, Wageningen University, Wageningen, The Netherlands) for supplying *S. exigua* eggs and to S. Reynolds and A. Meredith (University of Bath, Bath, United Kingdom) for *M. sexta* eggs.

N. Staykov was supported by a grant from the EC-TEMPUS program.

REFERENCES

1. **Aronson, A. I., D. Wu, and C. Zhang.** 1995. Mutagenesis of specificity and toxicity regions of a *Bacillus thuringiensis* protoxin gene. J. Biol. Chem. 177: 4059–4065.
2. **Bosch, D., B. Schipper, H. van der Kleij, R. de Maagd, and W. Stiekema.** 1994. Recombinant *Bacillus thuringiensis* crystal proteins with new properties: possibilities for resistance management. Bio/Technology 12:915–918.
3. **Burton, S. L., D. J. Ellar, J. Li, and D. J. Derbyshire.** 1999. N-acetylgalactosamine on the putative insect receptor aminopeptidase N is recognised by a site on the domain III lectin-like fold of a Bacillus thuringiensis insecticidal toxin. J. Mol. Biol. 287:1011–1022.
4. **de Maagd, R. A., P. L. Bakker, L. Masson, M. J. Adang, S. Sangadala, W. Stiekema, and D. Bosch.** 1999. Domain III of the *Bacillus thuringiensis* delta-endotoxin Cry1Ac is involved in binding to Manduca sexta brush border membranes and to its purified aminopeptidase N. Mol. Microbiol. 31: 463–471.
5. **de Maagd, R. A., M. S. G. Kwa, H. van der Klei, T. Yamamoto, B. Schipper, J. M. Vlak, W. J. Stiekema, and D. Bosch.** 1996. Domain III substitution in *Bacillus thuringiensis* delta-endotoxin Cry1A(b) results in superior toxicity for *Spodoptera exigua* and altered membrane protein recognition. Appl. Environ. Microbiol. 62:1537–1543.
6. **de Maagd, R. A., H. van der Klei, P. L. Bakker, W. J. Stiekema, and D. Bosch.** 1996. Different domains of *Bacillus thuringiensis* δ-endotoxins can bind to insect midgut membrane proteins on ligand blots. Appl. Environ. Microbiol. 62:2753–2757.
7. **Ge, A. Z., D. Rivers, R. Milne, and D. H. Dean.** 1991. Functional domains of *Bacillus thuringiensis* insecticidal crystal proteins. J. Biol. Chem. 266:17954–17958.
8. **Grochulski, P., L. Masson, S. Borisova, M. Pusztai-Carey, J. L. Schwartz, R. Brousseau, and M. Cygler.** 1995. *Bacillus thuringiensis* Cry1A(a) insecticidal toxin—crystal structure and channel formation. J. Mol. Biol. 254:447–464.
9. **Lee, M. K., B. A. Young, and D. H. Dean.** 1995. Domain III exchanges of *Bacillus thuringiensis* Cry1A toxins affect binding to different gypsy moth midgut receptors. Biochem. Biophys. Res. Commun. 216:306–312.
10. **Li, J. D., J. Carroll, and D. J. Ellar.** 1991. Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. Nature 353: 815–821.
11. **Masson, L., A. Mazza, L. Gringorten, D. Baines, V. Aneliunas, and R. Brousseau.** 1994. Specificity domain localization of *Bacillus thuringiensis* insecticidal toxins is highly dependent on the bioassay system. Mol. Microbiol. 14:851–860.
12. **Russel, R. M., J. L. Robertson, and N. E. Savin.** 1977. POLO: a new computer program for Probit analysis. ESA Bull. 23:209–213.
13. **Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. Baum, J. Feitelson, D. R. Zeigler, and D. H. Dean.** 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62:775–806.
14. **Van Rie, J., S. Jansens, H. Höfte, D. Degheele, and H. Van Mellaert.** 1990. Receptors on the brush border membrane of the insect midgut as determinants of the specificity of *Bacillus thuringiensis* delta-endotoxins. Appl. Environ. Microbiol. 56:1378–1385.

Journal of Biochemistry and Molecular Biology, Vol. 34, No. 5, September 2001, pp. 402-407

JOURNAL OF
Biochemistry
and
Molecular Biology

Directed Mutagenesis of the *Bacillus thuringiensis* Cry11A Toxin Reveals a Crucial Role in Larvicidal Activity of Arginine-136 in Helix 4

Chanan Angsuthanasombat*, Siriporn Keeratichamreon, Somphob Leetacheewa, Gerd Katzenmeier and Sakol Panyim

Laboratory of Molecular Biophysics, Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Nakompathom 73170, Thailand

Received 10 May 2001, Accepted 6 June 2001

Based on the currently proposed toxicity model for the different *Bacillus thuringiensis* Cry δ-endotoxins, their pore-forming activity involves the insertion of the α4-α5 helical hairpin into the membrane of the target midgut epithelial cell. In this study, a number of polar or charged residues in helix 4 within domain I of the 65-kDa dipteran-active Cry11A toxin, Lys-123, Tyr-125, Asn-128, Ser-130, Gln-135, Arg-136, Gln-139 and Glu-141, were initially substituted with alanine by using PCR-based directed mutagenesis. All mutant toxins were expressed as cytoplasmic inclusions in *Escherichia coli* upon induction with IPTG. Similar to the wild-type protoxin inclusion, the solubility of each mutant inclusion in the carbonate buffer, pH 9.0, was relatively low. When *E. coli* cells, expressing each of the mutant proteins, were tested for toxicity against *Aedes aegypti* mosquito-larvae, toxicity was completely abolished for the alanine substitution of arginine at position 136. However, mutations at the other positions still retained a high level of larvicidal activity. Interestingly, further analysis of this critical arginine residue by specific mutagenesis showed that conversions of arginine-136 to aspartate, glutamine, or even to the most conserved residue lysine, also abolished the wild-type activity. The results of this study revealed an important determinant in toxin function for the positively charged side chain of arginine-136 in helix 4 of the Cry11A toxin.

Keywords: *Bacillus thuringiensis*, δ-endotoxin, Inclusion solubility, Larvicidal activity, Site-directed mutagenesis

*To whom correspondence should be addressed.
Tel: 662-441-9003, 1278; Fax: 662-441-9906
E-mail: stcas@mahidol.ac.th

Introduction

Bacillus thuringiensis (*Bt*), a Gram-positive endospore-forming bacterium, produces insecticidal proteins in large quantities as different forms of parasporal crystalline inclusions during sporulation (Hofte and Whiteley, 1989). These cytoplasmic inclusions are composed of one or several polypeptides that have been classified as Cry and/or Cyt δ-endotoxins on the basis of the similarity of their deduced amino acid sequences (Hofte and Whiteley, 1989; Crickmore *et al.*, 1998). Currently, the Cry δ-endotoxins have been shown to be active against insect larvae in the orders Diptera (mosquitoes and flies), Lepidoptera (moths and butterflies), Coleoptera (beetles and weevils), and Hymenoptera (wasps and bees) (Schnepf *et al.*, 1998; de Maagd *et al.*, 2001). For instance, the 65-kDa Cry11A toxin and the 130-kDa Cry4B toxin that are produced from *Bt* subsp. *israelensis* are specifically toxic to mosquito larvae (Hofte and Whiteley, 1989; Schnepf *et al.*, 1998).

The *Bt* δ-endotoxins exist as inactive protoxins found within inclusion bodies, which require alkaline solubilisation and proteolytic activation in the insect larval midgut (Hofte and Whiteley, 1989). It has been proposed that, after activation by gut proteases, the active toxins kill the susceptible larvae *via* a two-step receptor mediated mechanism, in which the initial toxin-receptor interaction is followed by membrane insertion of the toxins to form transmembrane leakage pores. These pores cause the target midgut epithelial cells to swell and lyse by colloid-osmotic lysis (Knowles and Ellar, 1987), resulting in extensive damage to the midgut and eventually larval death (Knowles, 1994). However, the precise mechanism of action of the *Bt* toxins is still not completely understood, although knowledge of how these insecticidal proteins work at the molecular level has increased substantially over the last decade.

To date, the three-dimensional structures of two different Cry δ-endotoxins, Cry1Aa, and Cry3A have been solved by

X-ray crystallography (Li, Carroll and Ellar, 1991; Grochulski *et al.*, 1995). Both structures display a high degree of overall structural similarity and are composed of three structurally distinct domains. It is apparent that the N-terminal domain, a seven-helix bundle (six amphipathic helices around a central core helix), is clearly equipped for membrane insertion and pore formation (Li, Carroll and Ellar, 1991; Grochulski *et al.*, 1995). This suggestion has been supported by various studies demonstrating that the isolated helical fragment from different Cry toxins is responsible for pore-forming activity (Walters *et al.*, 1993; Von Tersch *et al.*, 1994; Puntheeranurak *et al.*, 2001).

The molecular mechanism of membrane insertion and pore formation of the Cry toxins is now described in an 'umbrella' model (Knowles, 1994). In this model, α4 and α5 form a helical hairpin to initiate membrane penetration upon specific receptor-binding in which structural rearrangement of the toxin occurs. After insertion of this hairpin, the other helices spread over the membrane surface followed by oligomerization of the toxin (Gazit *et al.*, 1998; Guereca and Bravo, 1999), resulting in formation of an initial tetrameric pore (Schwartz *et al.*, 1997). Currently, this model is supported by a number of experiments, which demonstrates the crucial role of α4 and α5 in pore-forming activity of different Cry toxins (Schwartz *et al.*, 1997; Kumar and Aronson, 1999; Masson *et al.*, 1999; Nunes-Valdes *et al.*, 2001). Recent studies clearly demonstrated that the helix 4-loop-helix 5 hairpin is more active in membrane penetration than each of the isolated helices, or their mixtures, consistent with its function as the membrane-inserted portion of the Cry toxins (Gerber and Shai, 2000).

In earlier studies, we demonstrated that α4 and α5 of the 130-kDa Cry4B toxin are essential determinants of toxicity, likely to be involved in pore formation rather than in receptor recognition (Uawithya *et al.*, 1998; Sramala *et al.*, 2000). In addition, arginine-158 in helix 4 was found to play an important role in larvicidal activity of this toxin (Sramala *et al.*, 2001). In the present report, an analogous effect on toxicity was observed for the 65-kDa dipteran-specific Cry11A toxin when charged, and polar residues in helix 4 were altered. The results revealed that the specific structure for the positively charged side chain of arginine-136 in this helix is directly involved in Cry11A toxin activity, supporting the notion that α4 is essential for pore formation by the Cry δ-endotoxins.

Materials and Methods

Plamids and site-directed mutagenesis The full-length gene, encoding the 65-kDa Cry11A toxin from the recombinant plasmid pBTC68A (a generous gift of Dr. Wattanalai Panbangred, Department of Biotechnology, Mahidol University, Thailand), was subcloned into the pMEx8 expression vector (Buttcher *et al.*, 1990). This resultant plasmid (pME4D) was used as a template for site-directed mutagenesis. Each complementary pair of mutagenic oligonucleotide primers was purchased from Genset Inc. (Singapore), as shown in Table 1. All mutations were generated by PCR using high fidelity *Pfu* DNA polymerase following the procedure of the QuickChange™ mutagenesis kit (Stratagene). All mutant plasmids were analyzed by DNA sequencing using a Perkin Elmer ABI prism 377 automated sequencer.

Table 1. Complementary primers for substituting a coded residue with different amino acids.

Primer	Sequence[a]	Restriction Site
	A T A A G Y F L N	
K123A-f	5'CTGCAACAG CTGCAGGTTATTTTCTAAATC 3'	*Psi*I
K123A-r	5'GATTTAGAAAATAAC CTGCAGCTGTTGCAG 3'	
	G A I I Q A L P Q F	
R136A-f	5'GTGGTGCTATAATAC AAGCTTTACCTCAATTTG 3'	*Hind*III
R136A-r	5'CAAATTGAGGTA AAGCTTGTATTATAGCACCAC 3'	
	I Q R L P Q F A V Q T	
E141A-f	5'TAATACAAC GTCTACCTCAATTTGCAGTTCAAACAT 3'	*Acc*I
E141A-r	5'ATGTTTGAAC TGCAAATTGAGGTAGACGTTGTATTA 3'	
	L N L S G A I I Q K L P	
R136K-f	5'CTAAATCTAAGT GGGGCCATAATACAAAAATTACCTC 3'	*Hae*III
R136K-r	5'GAGGTAATTTTTGTATTAT GGCCCCACTTAGATTTAG 3'	
	L N L S G A I I Q Q L P	
R136Q-f	5'CTAAATCTAAGT GGTGCCATAATACAACAATTACCTC 3'	*Nla*IV
R136Q-r	5'GAGGTAATTGTTGTATTAT GGCACCACTTAGATTTAG 3'	
	G A I I Q D L P Q F E	
R136D-f	5'GGTGCTATAATACA AGATCTACCTCAATTTGAGG 3'	*Bgl*II
R136D-r	5'CCTCAAATTGAGGT AGATCTTGTATTATAGCACC 3'	
	L S G A I I A R L P Q F	
Q135A-f	5'CTAAGT GGTGCCATAATAGCTAGGTTACCTCAATTTG 3'	*Bst*EII & *Nla*IV
Q135A-r	5'CAAATTGA GGTAACCTAGCTATTATGGCACCACTTAG 3'	
	A I I Q R L P A F E V	
Q139A-f	5'GCTATAATACAAC GTCTACCTGCTTTTGAGGTTC 3'	*Acc*I
Q139A-r	5'GAACCTCAAA AGCAGGTAGACGTTGTATTATAGC 3'	
	T A K G A F L N L	
Y125A-f	5'CAAC AGCCAAGGGTGCCTTTCTAAATCTAAG 3'	*Nla*IV
Y125A-r	5'CTTAGATTTAGAAA GGCACCCTTGGCTGTTG 3'	
	G Y F L A L S G	
N128A-f	5'GGGTTATTTTCTA GCATTAAGTGGTGC 3'	*Dde*I
N128A-r	5'GCACCACTTAATGCTAGAAAATAACCC 3'	
	Y F L N L A G A I I	
S130A-f	5'GTTATTTTCTAAATCT GGCCGGTGCTATAATAC 3'	*Hae*III
S130A-r	5'GTATTATAGCACC GGCCAGATTTAGAAAATAAC 3'	

[a]Introduced restriction enzyme recognition sites are underlined. The mutated nucleotide residues are shown as boldface. Deduced amino acid sequences are shown on top of each pair of oligonucleotide primers.

Toxin expression and characterization The wild type and mutant Cry11A toxin genes were expressed in the *E.coli* strain JM109 under control of the inducible *tac* promoter. Cells were grown in a LB medium that was supplemented with 100 μg ml^{-1} ampicillin until OD_{600} reached 0.4-0.5. Incubation was continued for another 4 h after addition of isopropyl-β-D-thiogalacto-pyranoside (IPTG) to the final concentration of 0.1 mM. Protein expression was analyzed by sodium dodecyl sulfate (SDS)-15% w/v polyacrylamide gel electrophoresis (PAGE). Immunoblotting was performed with polyclonal rabbit antibodies against the Cry11A toxin (kindly provided by Prof. David Ellar, University of Cambridge, UK). Immunocomplexes were detected with an anti-rabbit antibody-alkaline phosphatase conjugate (Sigma).

E. coli cultures, expressing each mutant as cytoplasmic inclusion bodies, were harvested by centrifugation, resuspended in distilled

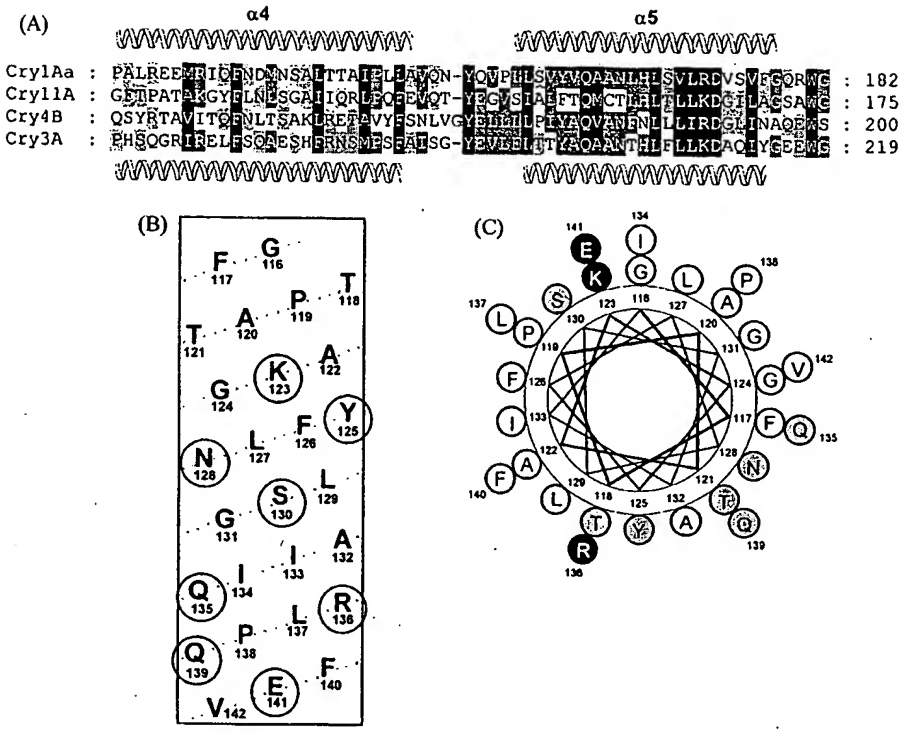


Fig. 1. (A) Multiple sequence alignment of helices 4 and 5 of Cry11A with the crystal structures of Cry1Aa and Cry3A, and the homology-based model of Cry4B. The sequences were aligned using the program CLUSTAL W. The degree of conservation is represented by background shading of the residues with the darkest shading for the most conserved: 100% conserved, 75% conserved, and 50% conserved. The positions of secondary structure elements of Cry1Aa and Cry3A are illustrated over and under the alignments, respectively. (B) The predicted pattern of helix 4 of Cry11A is composed of 27 residues of which the encircled residues were mutated. (C) A helical wheel projection of helix 4 of Cry11A. Amino acid residues are plotted every 100 degrees consecutively around the wheel, following the sequences given in B. The following color code is used: black is an amino acid with a charged side chain, gray is a polar side chain, and white is a hydrophobic side chain.

water, and disrupted in a French Pressure Cell at 16,000 psi. The crude lysates were centrifuged at 8,000 g for 5 min and the pellets obtained were washed 3 times in distilled water. Protein concentrations were determined by using a protein microassay (Bio-Rad) with the bovine serum albumin fraction V (Sigma) as a standard. Protoxin inclusions (1 mg ml^{-1}) were solubilized in 50 mM Na_2CO_3, pH 9.0 and incubated at 37°C for 60 min, as described previously (Uawithya *et al.*, 1998). After centrifugation for 10 min, the supernatants were analyzed by SDS-PAGE in comparison with the inclusion suspension.

Larvicidal activity assays Bioassays were performed, as described previously (Angsuthanasombat *et al.*, 1993), using 2-day old *Aedes aegypti* mosquitó-larvae reared from eggs that were supplied by the mosquito-rearing facility of the Institute of Molecular Biology and Genetics, Mahidol University, Thailand. About 500 larvae were reared in a container (22 × 30 × 10 cm) with approximately 3 litres of distilled water that was supplemented with a small amount of rat diet pellets. Both rearing and bioassays were performed at room temperature (25°C). The assays were carried out in 1 ml of *E. coli* suspension (10^8 cells suspended in distilled water) in a 48-well micrometer plate (11.3 mm well diameter) with 10 larvae per well and a total of 100 larvae for each type of *E. coli* samples. *E. coli* cells, containing the recombinant plasmid pME4D and the pMEx8 vector, were used as positive and negative controls, respectively. Mortality was recorded after a 24-hour incubation period.

Results and Discussion

Based on a multiple-amino acid sequence alignment with the known crystal structures of Cry1Aa and Cry3A (Li, Carroll, and Ellar, 1991; Grochulski *et al.*, 1995), and the homology-based model of Cry4B (Uawithya *et al.*, 1999), the predicted α4 and α5 were located within the pore-forming domain of Cry11A (see Fig. 1A). Charged amino acids in helix 4 were shown to be critical for toxin activity (Kumar and Aronson, 1999; Masson *et al.*, 1999; Sramala *et al.*, 2001). To investigate the possible role for toxicity of charged and polar amino acids in α4 of Cry11A, a PCR-based mutagenesis strategy, previously employed for Cry4B (Sramala *et al.*, 2001), was applied to obtain substitutions within Cry11A. We initially generated eight Cry11A mutants in which three

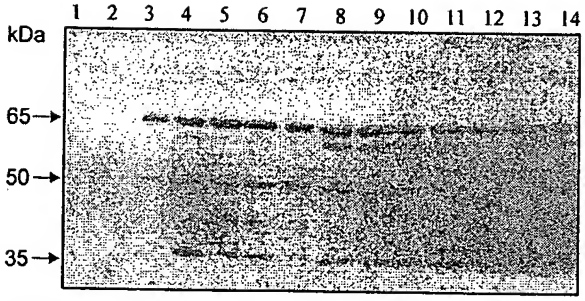


Fig. 2. Western blot analysis of lysates extracted from *E. coli* cells harboring the pMEx8 vector (lane 2), pME4D (lane 3), or the mutant plasmids-K123A, Y125A, N128A, S130A, Q135A, R136A, R136K, R136Q, R136D, Q139A, and E141A (lanes 4-14, respectively)-showing the 65-kDa Cry11 toxin and small molecular fragments that cross-reacted with the Cry11A antibodies. Lane 1 represents the molecular mass standards.

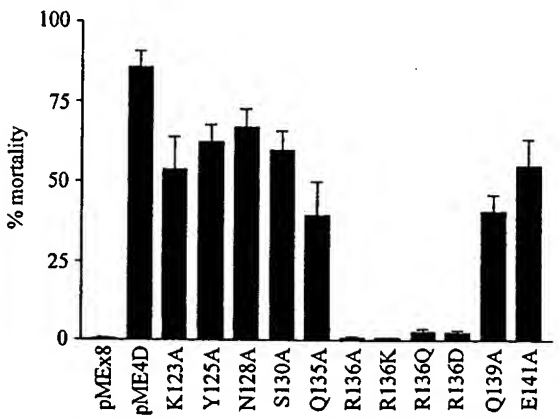


Fig. 3. Mosquito-larvicidal activities of *E. coli* cells expressing the Cry11A wild-type toxin (pME4D), or its mutants-K123A, Y125A, N128A, S130A, Q135A, R136A, R136K, R136Q, R136D, Q139A and E141A)-against *Aedes aegypti* larvae. Error bars indicate standard error of the mean from the three independent experiments.

charged and five polar amino acids in helix 4 (Fig. 1B) were substituted with alanine. Most of the targeted residues, Tyr-125, Asn-128, Gln-135, Arg-136 and Gln-139, but not Lys-123, Ser-130 and Glu-141, are located at the hydrophilic surface (see Fig. 1C).

Expression of the mutant toxins in *E. coli* was controlled by the *tac* promoter. Upon addition of IPTG to mid-exponential phase cultures, all mutant toxins were predominantly produced as sedimentable inclusion bodies. Lysates were analyzed by SDS-PAGE and immunoblotting, and the protein expression level of all mutant derivatives was found to be comparable to the wild type. The 65-kDa expressed mutant proteins specifically cross-reacted with antibodies raised against the Cry11A toxin (see Fig. 2). However, two relatively intense immuno-reactive bands of ca. 50 kDa and ca. 35 kDa were detected in all mutant lysates. This indicates that the expressed mutant proteins are rather sensitive to proteolytic degradation.

The solubility of mutant protein inclusions in comparison to the wild-type inclusion was assessed using a carbonate buffer, pH 9.0. The amount of 65-kDa soluble proteins in the supernatant was compared with those of the proteins initially used, in order to determine the percentage of protein solubilisation. All of the mutant inclusions were found to be soluble to some extent in this buffer, giving less than 20% solubility, which resembles closely the wild-type inclusions under similar conditions.

To determine the effect of mutations on toxicity, *E. coli* cells that expressed each type of the mutant toxin were tested for their relative biological activity towards *Aedes aegypti* larvae. All of the assays were carried out in ten replicas for each sample and repeated three times; the mortality data recorded after a 24-hour incubation are shown in Fig. 3. Interestingly, only the R136A mutation resulted in a total loss of larvicidal activity, while alanine substitutions at seven other positions (K123A, Y125A, N128A, S130A, Q135A, Q139A and E141A) still retained over 50% of the wild-type activity. When this critical arginine residue at position 136 was converted to aspartate, glutamine or even to the most conserved residue for positively charged side chain, *i.e.* lysine, all R136 mutants (R136D, R136Q and R136K) were shown to be nontoxic to mosquito larvae (see Fig. 3). These results could imply the requirement for a specific structure of the positive side chain at this position. Perhaps Arg-136, which is likely to face the pore lumen, could interact with an aqueous environment, and somehow stabilize the functional pore. However, the precise function of this residue remains to be elucidated.

Protein expression levels and solubility of the inclusions suggested that the complete loss of toxicity observed for the R136A mutant is least likely to be caused by misfolding of the protein. Taken together, our results indicate that Arg-136 is a critical residue involved in Cry11A toxin activity. The data further support our previous findings that Arg-158 in α4 played a crucial role in toxicity of the 130-kDa Cry4B toxin, since the single alanine substitution at this residue almost completely abolished its activity towards mosquito larvae (Sramala *et al.*, 2001). In addition, results reported by other workers revealed that an arginine residue at position 131 in α4 is important for toxicity of both the lepidopteran-specific Cry1Aa and Cry1Ac toxins (Kumar and Aronson, 1999; Masson *et al.*, 1999). Two other negatively charged residues (Glu-129 and Asp-136) of Cry1Aa were also shown to be critical in the passage of ions through the pore (Masson *et al.*, 1999).

Comparisons of structural models among Cry11A, Cry4B, and Cry1Aa suggest that, although Arg-136 of Cry11A is located on the opposite side of helix 4 relative to Arg-158 of Cry4B or Arg-131 of Cry1Aa, all of these three critical

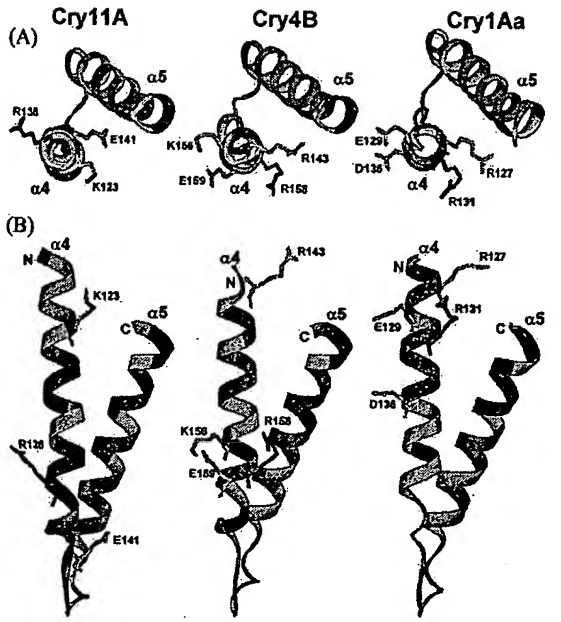


Fig. 4. (A) Top and (B) side views of amino acid arrangement in helix 4 together with the relative position of helix 5 in 3D models of Cryl1A and Cry4B built by homology modeling and the Cry1Aa crystal structure. The labeled residues, shown in red and yellow, indicate the critical arginine residues and other charged positions, respectively. The structures were prepared using Weblab viewer (Molecular Simulations Inc.).

residues are oriented on the side of helix 4, which is furthest away from helix 5 (see Fig. 4A). It should be noted that Arg-136 and Arg-158, in both of the dipteran-specific toxins, are situated near the C-terminal end of helix 4, while Arg-131 of the lepidopteran-specific Cry1Aa toxin is located furthest from the C-terminus of this helix (Fig. 4B). Differences in the location of these critical residues may conceivably reflect the diversity in the channel architecture for each group of insect-specific Cry toxins. Further studies are required to elucidate the role of these positively charged residues in helix 4, to discover whether they are involved in the passage of ions through the pore.

Acknowledgments We are grateful to Prof. Prapon Wilairat for the helpful comments. We thank U. Chanama for the excellent technical support with DNA sequencing, as well as A. Nirachanon and S. Shimwai for technical assistance. This work was supported in part by the National Science and Technology Development Agency of Thailand and the Thailand Research Fund.

References

Angsuthanasombat, C., Crickmore, N. and Ellar, D. J. (1993) Effects on toxicity of eliminating a cleavage site in a predicted interhelical loop in *Bacillus thuringiensis* CrylVB delta-endotoxin. *FEMS Microbiol. Lett.* **111**, 255-261.

Buttcher, V., Ruhlmann, A. and Cramer, F. (1990) Improved single-stranded DNA producing expression vectors for protein manipulation in *Escherichia coli*. *Nucleic Acids Res.* **18**, 1075.

Crickmore, N., Zeigler, D. R., Feitelson, J., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J. and Dean, D. H. (1998) Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **62**, 807-813.

de Maagd, R. A., Bravo, A. and Crickmore, N. (2001) How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. *Trends Genet.* **4**, 193-199.

Gazit, E., La Rocca, P., Sansom, M. S. and Shai, Y. (1998) The structure and organization within the membrane of the helices composing the pore-forming domain of *Bacillus thuringiensis* delta-endotoxin are consistent with an "umbrella-like" structure of the pore. *Proc. Natl. Acad. Sci. USA* **95**, 12289-12294.

Gerber, D. and Shai, Y. (2000) Insertion and organization within membranes of the delta-endotoxin pore-forming domain, helix 4-loop-helix 5, and inhibition of its activity by a mutant helix 4 peptide. *J. Biol. Chem.* **275**, 23602-23607.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J. L., Brousseau, R. and Cygler, M. (1995) *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. *J. Mol. Biol.* **254**, 447-464.

Guereca, L. and Bravo, A. (1999) The oligomeric state of *Bacillus thuringiensis* Cry toxins in solution. *Biochem. Biophys. Acta* **1429**, 342-350.

Hofte, H. and Whiteley, H. R. (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Knowles, B. H. and Ellar, D. J. (1987) Colloid-osmotic lysis is a general feature the mechanism of action of *Bacillus thuringiensis* δ-endotoxins with different insect specificity. *Biochem. Biophys. Acta* **924**, 509-518.

Knowles, B. H. (1994) Mechanism of action of *Bacillus thuringiensis* insecticidal δ-endotoxins. *Adv. Insect. Physiol.* **24**, 275-308.

Kumar, A. S. and Aronson, A. I. (1999) Analysis of mutations in the pore-forming region essential for insecticidal activity of a *Bacillus thuringiensis* delta-endotoxin. *J. Bacteriol.* **181**, 6103-6107.

Li, J. D., Carroll, J. and Ellar, D. J. (1991) Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. *Nature* **353**, 815-821.

Masson, L., Tabashnik, B. E., Liu, Y. B., Brousseau, R. and Schwartz, J. L. (1999) Helix 4 of the *Bacillus thuringiensis* Cry1Aa toxin lines the lumen of the ion channel. *J. Biol. Chem.* **274**, 31996-32000.

Nunez-Valdez, M.-E., Sanchez, J., Lina, L., Guereca, L. and Bravo, A. (2001) Structural and functional studies of α-helix 5 region from *Bacillus thuringiensis* Cry1Ab δ-endotoxins. *Biochimi. Biophys. Acta* **1546**, 122-133.

Puntheeranurak, T., Leetacheewa, S., Katzenmeier, G., Krittanai, C., Panyim, S. and Angsuthanasombat, C. (2001) Expression and biochemical characterization of the *Bacillus thuringiensis* Cry4B α1-α5 pore-forming fragment. *J. Biochem. Mol. Biol.* (in press).

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R. and Dean, D. H. (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol.*

Biol. Rev. **62**, 775-806.

Schwartz, J. L., Juteau, M., Grochulski, P., Cygler, M., Prefontaine, G., Brousseau, R. and Masson, L. (1997) Restriction of intramolecular movements within the Cry1Aa toxin molecule of *Bacillus thuringiensis* through disulfide bond engineering. *FEBS Lett.* **410**, 397-402.

Sramala, I., Uawithya, P., Chanama, U., Leetachewa, S., Krittanai, C., Katzenmeier, G., Panyim, S. and Angsuthanasombat, C. (2000) Single proline substitutions of selected helices of the *Bacillus thuringiensis* Cry4B toxin affect inclusion solubility and larvicidal Activity. *J. Biochem. Mol. Biol. Biophys.* **4**, 187-193.

Sramala, I., Leetachewa, S., Krittanai, C., Katzenmeier, G., Panyim, S. and Angsuthanasombat, C. (2001) Screening charged residues in helix 4 of the *Bacillus thuringiensis* Cry4B toxin reveals one critical residue for larvicidal activity. *J. Biochem. Mol. Biol. Biophys.* (in press).

Uawithya, P., Krittanai, C., Katzenmeier, G., Leetachewa, S., Panyim, S. and Angsuthanasombat, C. (1999) 3D models for *Bacillus thuringiensis* Cry4A and Cry4B insecticidal proteins based on homology modeling. Abstract 32nd Meeting, Society for Invertebrate Pathology, Irvine, USA, p. 75.

Uawithya, P., Tuntitippawan, T., Katzenmeier, G., Panyim, S. and Angsuthanasombat, C. (1998) Effects on larvicidal activity of single proline substitutions in α3 or α4 of the *Bacillus thuringiensis* Cry4B toxin. *Biochem. Mol. Biol. Int.* **44**, 825-832.

Von Tersch, M. A., Slatin, S. L., Kulesza, C. A. and English, L. H. (1994) Membrane-permeabilizing activities of *Bacillus thuringiensis* coleopteran-active toxin CryIIIB2 and CryIIIB2 domain I peptide. *Appl. Environ. Microbiol.* **60**, 3711-3717.

Walters, F. S., Slatin, S. L., Kulesza, C. A. and English, L. H. (1993) Ion channel activity of N-terminal fragments from CryIA(c) delta-endotoxin. *Biochem. Biophys. Res. Commun.* **196**, 921-926.

Attorney Docket No. 035718/237005 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re:	Abad *et al.*	Confirmation No.:	5409
Appl. No.:	10/032,717	Group Art Unit:	1638
Filed:	10/23/2001	Examiner:	A.R. Kubelik
For:	GENES ENCODING NOVEL BACILLUS THURINGIENSIS PROTEINS WITH PESTICIDAL ACTIVITY AGAINST COLEOPTERANS		

January 18, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RULE 132 DECLARATION
of
André Abad

Sir:

I, André Abad, do hereby declare and say as follows:

I am skilled in the art of the field of the invention of the above-referenced application. I earned the following academic degrees: BS majoring in mathematics and biochemistry from the Wisconsin University River Falls in 1978 and Ph.D. from Purdue University Department of Agronomy in 1996. My thesis investigated the role of a mitochondrial gene in cytoplasm male sterility in beans. From 1979 to 1991, I was employed by the University of Minnesota, Department of Plant Pathology. Working in Dr. Blanchette's laboratory, we investigated and published numerous manuscripts in the area of woody tissue degradation by fungi, in particular related to the degradation of cell wall component by fungal enzymes such as xylanase. From 1996 to 1998, I worked in Dr. Judy Bond's laboratory at Hershey Medical Center in Pennsylvania. I was involved in characterization of mouse meprin receptors and generated the constructs and ES cells necessary for producing transgenic mice targeting the knockout of meprin. Since 1999, Pioneer Hi-Bred has employed me. My current responsibility as a research scientist is to

lead a research team for insecticidal protein optimization and genomic screening of *Bacillus thuringiensis* DNA for novel insecticidal genes

1. I am familiar with the experiments described in the above-mentioned application. Particularly, the procedures described in Examples 4, 6, and 7 of the above-mentioned application are considered "routine" by scientists who are familiar with research on endotoxins. Moreover, the production of plants expressing proteins having pesticidal activity, while it is a time-consuming and laborious task, is also considered "routine" by scientists who are responsible for producing such plants.

2. As one of skill in the art, given the disclosure in the above-referenced application, I would be able to make and use the claimed invention. For example, I would be able to make and use the nucleic acid of claim 1 by generating a collection of nucleic acids comprising a nucleotide sequence meeting the sequence limitation of the claims (*i.e.*, a nucleotide sequence having at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1) and assaying the encoded polypeptide for defensive activity as described in the specification, for example, as described in any of Examples 4, 6, and 7. Further, where such a collection included dozens of such sequences, I would consider this degree of experimentation to be routine rather than to be "undue experimentation." For these reasons, I believe that the claims, including for example, claim 1, are fully enabled and described by the specification.

3. It is my understanding, as one of skill in the art, that proteins can be produced that share a relatively low degree of sequence identity—maybe even as low as 70% sequence identity--with a known protein but that have the same or essentially the same function.

4. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 01-19-2005

By: _____
André Abad

42. Porter, N. A. & Weekes, T. C. *Mon. Not. R. astr. Soc.* 183, 205–210 (1978).
43. Akerlof, C. W. *et al.* in *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 135–138 (Adelaide University, 1990).
44. Porter, N. A. & Weekes, T. C. *Astrophys. J.* 212, 224–226 (1977).
45. Porter, N. A. & Weekes, T. C. *Mon. Not. R. Astr. Soc.* 183, 205–210 (1978).
46. Porter, N. A. & Weekes, T. C. *Nature* 277, 199 (1979).
47. Nolan, K., Porter, N. A., Fegan, D. J., Chantell, M. & Weekes, T. C. *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 150 (Adelaide University, 1990).
48. Fegan, D. J., McBreen, B., O'Brien, D. & O'Sullivan, C. *Nature* 271, 731–732 (1978).
49. Bhat, P. N. *et al. Nature* 284, 433–434 (1984).
50. Porter, N. A. & Weekes, T. C. *Nature* 267, 500–501 (1977).
51. O'Mongain, E. *Nature* 242, 136–137 (1973).
52. Huguenin, G. R. & Moore, E. L. *Astrophys. J.* 187, L57–L58 (1974).
53. Phinney, S. & Taylor, J. H. *Nature* 277, 117–118 (1979).
54. Rees, M. J. *Nature* 266, 333–334 (1977).
55. Jelley, J. V., Baird, G. A. & O'Mongain, E. *Nature* 267, 499–500 (1977).

ACKNOWLEDGEMENTS. We thank C. J. Goebel for giving his time whenever we needed expert help. We thank N. Porter for discussions. This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by the US Department of Energy, in part by the Xunta de Galicia (Spain) and in part by the Smithsonian Scholarly Studies Research Fund. J.H.M. is a NAS/NRC research associate.

ARTICLES

Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution

Jade Li*, Joe Carroll† & David J. Ellar†

* Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Biochemistry Department, Cambridge University, Tennis Court Road, Cambridge CB2 1QW, UK

The structure of the δ-endotoxin from *Bacillus thuringiensis* subsp. *tenebrionis* that is specifically toxic to Coleoptera insects (beetle toxin) has been determined at 2.5 Å resolution. It comprises three domains which are, from the N- to C-termini, a seven-helix bundle, a three-sheet domain, and a β sandwich. The core of the molecule encompassing all the domain interfaces is built from conserved sequence segments of the active δ-endotoxins. Therefore the structure represents the general fold of this family of insecticidal proteins. The bundle of long, hydrophobic and amphipathic helices is equipped for pore formation in the insect membrane, and regions of the three-sheet domain are probably responsible for receptor binding.

THE δ-endotoxins are a family of insecticidal proteins produced by *Bacillus thuringiensis* (B.t.) during sporulation, having relative molecular masses (M_r) 60,000–70,000 (60K–70K) in the active form and specific toxicities against insects in the orders of Lepidoptera, Diptera and Coleoptera[1,2]. These toxins have been formulated into commercial insecticides for three decades[3], and now insect-resistant plants are engineered by transformation with Lepidoptera-specific toxin genes[4–6]. In the bacterium δ-endotoxins are synthesized as protoxins of M_rs 70K–135K and crystallize as a parasporal inclusion ~1 μ in size, in which form they are ingested by the susceptible insect. The microcrystal dissolves in the alkaline pH of the midgut and the protoxin is cleaved by gut proteases to release the active toxin. δ-Endotoxins activated *in vitro* bind specifically and with high affinity ($k_D \approx$ 0.1–20 nM) to protein receptors on brush-border membrane vesicles derived from the gut epithelium of target insects[7–9] and create leakage channels of 10–20 Å diameter in the cell membrane[10]. *In vivo* such membrane lesions lead to swelling and lysis of the gut epithelium[11] and death of the insect ensues through starvation and septicaemia. Active δ-endotoxins of different specificities show five strongly conserved regions in their amino-acid sequences[1,12]. Exchanging sequence segments in the divergent regions between toxins of different specificities can produce active hybrids showing altered target specificity[13–15]. We have determined the atomic structure of a Coleoptera-specific δ-endotoxin (CryIIIA, beetle toxin) from *B.t.* subsp. *tenebrionis*[16–18] to elucidate the structural basis for target specificity and membrane perforation by this family of proteins.

Structure determination

Parasporal crystals of the beetle toxin contain the full-length 644-residue protoxin[17] as the minor component, and a product of bacterial processing with 57 residues removed from the N-terminus as the major component[19]. The latter (M_r 67K) is similar in sequence to the active form of other δ-endotoxins. After solubilization, papain cleavage converts the mixture to the 67K toxin (see legend to Table 1). This was recrystallized in the original crystal form of the parasporal crystals, space group $C222_1$ and cell dimensions 117.1 by 134.2 by 104.5 Å, containing one molecule per asymmetric unit and 55% solvent by volume[18].

Initial evaluation of derivatives was carried out at 4.5 Å resolution with data collected on the FAST TV diffractometer[20] using CuKα radiation. Complete datasets (Table 1) were then collected to 2.5 Å resolution from native crystals using the imaging plate systems at the EMBL outstation at DESY and from the mercury and platinum derivatives on film at SRS Daresbury. The electron density map (Fig. 1) at 2.5 Å resolution calculated with phases from multiple isomorphous replacement (mean figure of merit, 0.63) was easily interpretable and was improved by solvent flattening[21,22]. A continuous polypeptide chain from residue 61 to residue 644 at the C terminus was traced unambiguously, and most side-chain atoms could be located in the map. The atomic model was built using the graphics program O (ref. 23) and had an initial *R*-factor of 37% for all data to 2.5 Å. After preliminary refinement using the program X-PLOR (ref. 24), the current model, containing 584 amino acid residues and 40 bound water molecules, has an *R*-factor of 19.9% and r.m.s. bond length deviation of 0.017 Å.

Description of the structure

Overview. The beetle toxin is a wedge-shaped molecule with a radius of gyration of 58 Å. As shown in Fig. 2*a*, it comprises three domains. Domain I, from the N terminus of the 67K toxin to residue 290, is a seven-helix bundle in which a central helix is completely surrounded by six outer helices tilted at about +20° to it (Fig. 3*b*,*c*). Domain II, from residues 291 to 500, contains three antiparallel β sheets packed around a hydrophobic core with a triangular cross-section (Fig. 4). Domain III, from residues 501 to 644 at the C terminus is a sandwich of two antiparallel β sheets (Fig. 5). Domains I and III make up the

TABLE 1 Data collection and phasing statistics

Data collection

Data	Method of collection	Number of crystals	Resolution (Å)	Number of measurements	Unique reflections (% completeness)	R_{merge}
Native	image plate	8	2.5	121,767	27,727 (100)	0.108
CH_3HgNO_3	film	7	2.5	103,623	27,767 (100)	0.095
$Hg(CH_3COO)_2$	film	5	2.5	60,224	25,919 (94.5)	0.103
cis-$Pt(NH_3)_2Cl_2$	film	7	2.5	86,629	25,924 (94.5)	0.107
K_2OsO_4	FAST	1	4.5	21,143	4,680 (100)	0.077
$HoCl_3$	FAST	1	4.5	20,013	4,701 (100)	0.069

Phasing statistics

Derivative	Anomalous data	Number of sites	R_{deriv}†	R_{Cullis}‡	Phasing power§ (resolution, Å)
CH_3HgNO_3	no	3	0.183	0.715	1.56 (2.5)
$Hg(CH_3COO)_2$	yes	6	0.247	0.609	2.28 (2.5)
cis-$Pt(NH_3)_2Cl_2$	no	5	0.185	0.682	1.54 (2.5)
K_2OsO_4	no	4	0.149	0.757	1.26 (5.5)
$HoCl_3$	no	3	0.095	0.741	1.35 (5.0)

Protein preparation: Solubilized parasporal crystals from *B.t.* subsp *tenebrionis* were incubated at 0.5 mg ml^{-1} protein with 0.125 units per ml Agarose-linked papain (Boehringer) in 3.3 M NaBr, 0.05 M sodium phosphate, pH 7.0, and 0.1 mg ml^{-1} phenylmethylsulphonylfluoride (PMSF) for 30 min 20 °C. Digestion was stopped by adding tosyl lysinechloromethylketone (TLCK) to 0.125 mg m^{-1} and Na_2CO_3 to one fifth volume and removing t enzyme-beads. The 67K beetle toxin was then purified by gel filtration on Sephadex G75 equilibrated with 0.1 M $NaHCO_3$, pH 10.5, 0.5 M NaBr. **Crystallizati** Single crystals were obtained by microdialysis at a protein concentration of 2.5 mg ml^{-1} against 0.1 M $NaHCO_3$, pH 10.5, 0.5 M NaBr. **Crystallizatic** against 0.1 M $NaHCO_3$, pH 9.2, 0.5 M NaBr at 16 °C; 3 mM NaN_3, 0.1 mM PMSF and 0.1 mg ml^{-1} TLCK were present in all buffers. Crystals were transferr by stages to 0.05 M 2-(*N*-morpholino)ethanesulphonic acid (MES), pH 6.5, for derivative preparation and mounted in 0.03% low-melting agarose in this buff during data collection. **Data collection:** Image plate and film data were processed using MOSFLM (Imperial College, London) and CCP4 programs (Daresbu UK). FAST (ref. 20) data were collected and processed with MADNES[45], and scaled in 3° batches. **Derivatives:** Crystals were soaked respectively in 0.25 m CH_3HgNO_3 for 3.5 h, in 1 mM $Hg(CH_3COO)_2$ for 14 h, in freshly prepared 1 mM *cis*-$Pt(NH_3)_2Cl_2$ for 21 h, in saturated K_2OsO_4 for 35 h, and in 2 mM $HoCl_3$ f 3 days. **Phase calculation:** Two heavy-atom sites in each derivative were located from difference Patterson functions, except in the case of $Hg(CH_3COO$ for which 3 sites were located, and the remaining sites were found by cross-phased difference Fouriers. Heavy-atom parameters were refined again centric data and phases calculated for all data using the program PHARE (G. Bricogne). The two low-resolution derivatives were refined against phase calculated from the high-resolution derivatives. Phasing with the three high-resolution derivatives gave an overall figure of merit of 0.61 (25–2.5 Å) and clearly interpretable map. Including the remaining derivatives slightly improved the connectivity of the map (overall figure of merit 0.63), and four cycles solvent flattening using a 50% solvent content and a 9 Å radius in mask calculation[21,22] improved the overall definition of densities. The starting mod was built using the program *O* (ref. 23) with the Bones option for main-chain tracing and the autobuild and manip options for side chains. Refinement t simulated annealing using the program X-PLOR (ref. 24) reduced the *R*-factor from 0.37 to 0.25 without individual *B*-factors, and to 0.23 with restraine individual *B*-factors. The model was adjusted in the loops 154–156, 429–436, and 483–488, and had 40 solvent molecules added, then refined by X-PLO again. The current model has an *R*-factor of 19.9%, with r.m.s. bond length deviation of 0.017 Å, r.m.s. bond angle deviation of 3.2°, and average atom *B*-factor of 18 Å^2.

* $R_{merge} = \sum_i \sum_i |I_i - \langle I \rangle| / \sum_i |\langle I \rangle|$, where I_i are intensity measurements for a reflection, and $\langle I \rangle$ is the mean intensity for this reflection.

† $R_{deriv} = \sum |F_{PH} - F_P| / \sum |F_P|$, where F_{PH} is the structure factor amplitude of the derivative crystal and F_P is that of the native.

‡ $R_{Cullis} = \sum ||F_{PH} \pm F_P| - F_H(calc)| / \sum |F_{PH} - F_P|$, where F_P and F_{PH} are defined as for R_{deriv}, and $F_H(calc)$ is the calculated heavy-atom structure factor amplitud summed over centric data only.

§ Phasing power = $\langle F_H \rangle / E$, the r.m.s. heavy-atom structure factor amplitudes divided by the residual lack of closure error.

bulky end of the molecule. Through their contact one of the two β sheets in domain III is almost entirely buried. To our knowledge (see, for example, ref. 25), the packing of helices in domain I and of sheets in domain II are both novel arrangements.

Domain I. The central helix in this seven-helix bundle is α_5 (Fig. 3*b*,*c*), which is oriented with its C terminus towards the bulky end of the molecule. Viewed from this end, the outer helices are arranged anticlockwise in the order of α_1, α_2, α_3, α_4, α_6 and α_7, with helices α_1 and α_7 adjacent to the β-sheet domains; α_2 is interrupted by a non-helical section and only the leading half, α_{2a}, is packed against α_5. Figure 3*a* shows the alignment of amino-acid sequence on the surfaces of the helices. The helices are long, especially α_3 to α_7, which contain respectively 8, 7, 6, 9 and 7 complete helical turns and hence would be long enough to span the 30-Å thick hydrophobic region of a membrane bilayer. Furthermore, the six outer helices bear a strip of hydrophobic residues (defined by $\Delta G \geq 0$ for transfer from oil to water) down their entire length on the side-facing helix α_5, so they are amphipathic. In keeping with the general observation that secondary structures are close-packed and bury hydrophobic surfaces[26], the helix contact angles in this domain cluster around +20° rather than −50°, giving the bundle a bouquet-like appearance (Fig. 3*b*). Figure 3*c* shows the bundle in cross-section. The interhelical space contains 27 aromatic residues which are packed in the edge-to-face fashion[27]; all polar groups in this region are hydrogen-bonded or in salt bridges.

The concentric arrangement of the seven-helix bundle is dis tinct from the two-layered type seen in bacteriorhodopsin. Ther is some resemblance to the pore-forming domain of colicin A[28] in which two hydrophobic helices are shielded from solvent b eight amphiphilic helices, but the colicin helices are generall shorter. Like the colicin helices, the bundle in the beetle toxi may be a soluble form of packaging for the hydrophobic an amphiphilic helices that will form pores in the membrane afte a large change in conformation.

Domain II. In Fig. 4*a* and 4*b* the three sheets of this domain ar laid side-by-side, as they would be seen from the solvent. Ther is an apparent structural duplication between the four-strande antiparallel sheets, sheet 1 and sheet 2. The chain connections β_4, β_3, β_2, β_5 and β_8, β_7, β_6, β_9, respectively, follow the orde of +3, −1, −1, +3, which is typical of the 'Greek-key' topology[29] From both sheets the inner strands, β_3 and β_2 as well as β_7 and β_6, extend some 20 Å to the apex of the molecule as two-stranded β ribbons; and at the point of departure from the sheets there is a β-bulge in β_3 and in β_7 to twist the plane of the ribbon by nearly 90° relative to the sheet. The connections between the outer strands cross over the ribbons on the solvent side.

The pseudo-symmetry between these sheets is very approximate. Using the least squares option in *O* (ref. 23), the sheet region of the strands β_3 and β_2 can be brought to superimpose on that of β_7 and β_6, with a r.m.s. fit of 0.72 Å for 13 α carbons. But the r.m.s. fit increased to 1.1 Å for 23 α carbons of the

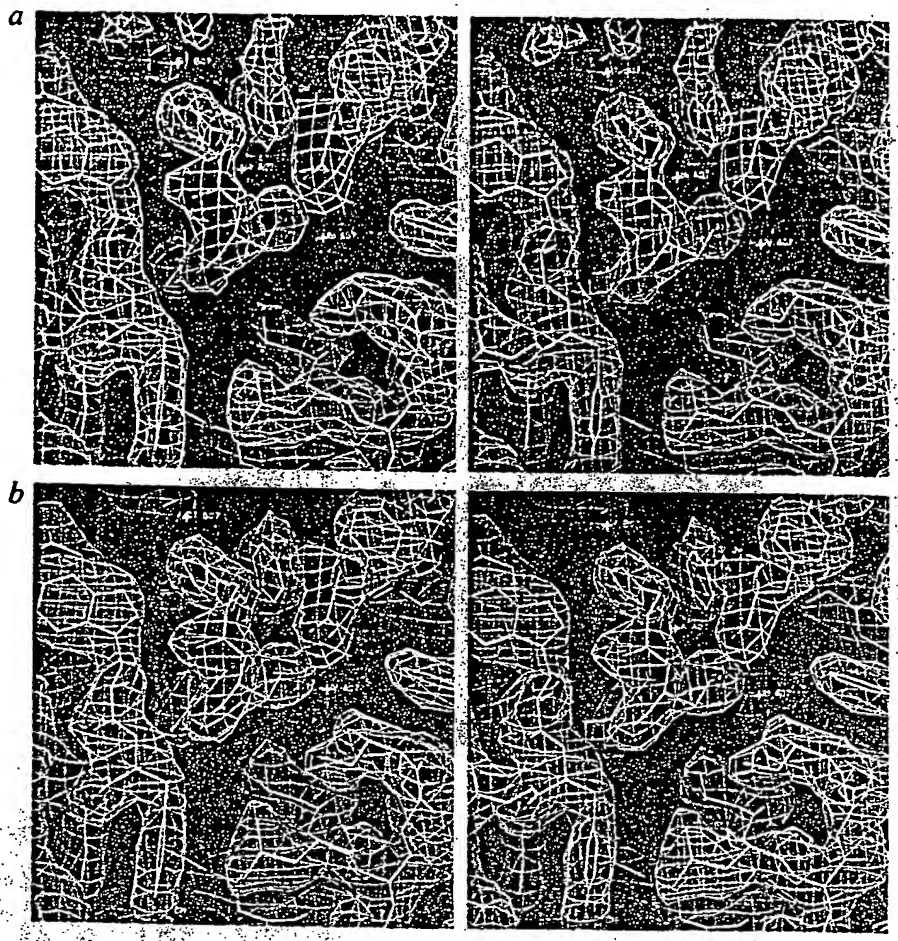


FIG. 1 Electron density map in the neighbourhood of Cys 243, calculated *a*, using combined phases[46] from multiple isomorphous replacement and solvent flattening, and *b*, using combined experimental and model phases[46] after refinement by X-PLOR. The refined structure is shown superimposed for reference. Although Cys 243 is a major site of both the methylmercury (MM) and mercuric acetate (MA) derivatives, the methyl mercury site is in a hydrophobic environment compared with the mercuric acetate site.

whole inner strands including the ribbon region, and 1.7 Å for 36 α carbons on all four strands. Nonetheless, the sequence alignment brought by this superposition of the two sheets revealed a low level of internal homology, with seven pairs of equivalent residues (shown in bold) out of 41 aligned α carbons:

```
338 HRIQPHTRPQP(6)SFNYWS(1)NYVSTRPSI(0)GSNDIITSPF(10)NLKFN 395
402 AVANTNLAVWP(0)SAVYSG(1)TKVEPSQYN(3)DEASTQTYDS(7 )SWDSI 453
```

The three-stranded sheet 3 is formed by two separate polypeptide segments. The C-terminal segment of domain II contributes the two-stranded ribbon of β_{10} and β_{11}, whereas the N-terminal segment of this domain contributes strand β_1, which is hydrogen-bonded to β_{11}; β_1 is followed by a two-turn helix α_8 and an extended chain.

Figure 4*c* and *d* shows in side view and in cross-section that the three antiparallel sheets are packed around a triangular hydrophobic core. This brings the strand β_{10} on the edge of sheet 3 into proximity with strand β_4 on the edge of sheet 1, as well as placing the loops at the end of the three β ribbons into a region of about 12 Å radius at the molecular apex. This domain is in contact with helix α_7 of domain I on the face of sheet 3 (Fig. 4*c*).

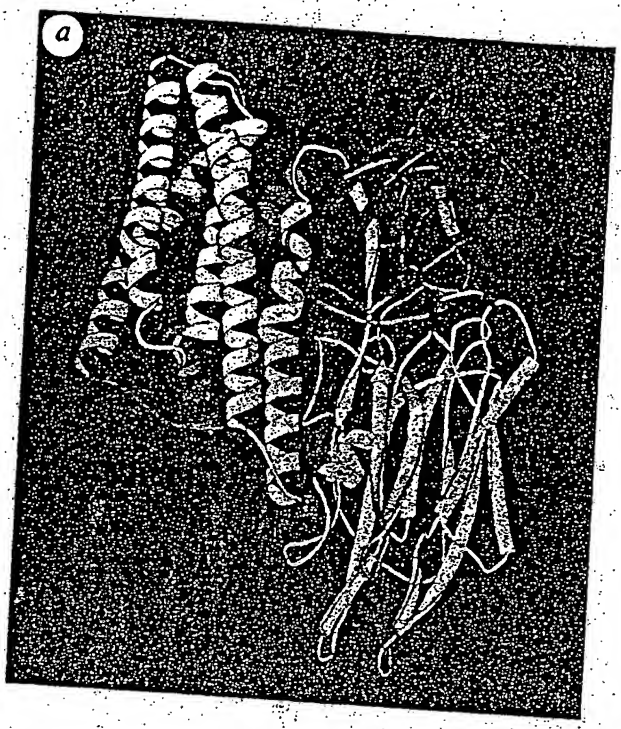
Domain III. Figure 5 is a ribbon drawing of the strands forming the two sheets of the β sandwich. The sheet containing the C-terminal strand is in contact with domain I and will be called the inner sheet. This domain has the 'jelly-roll' topology[29], because it can be generated by folding an antiparallel β ribbon which starts with β_{13} (N terminus) and β_{23} (C terminus) on the inner sheet, and ends in the loop between β_{18} and β_{19} on the outer sheet; β_{14} is a short excursion from this ribbon and forms the fifth antiparallel strand of the outer sheet. In addition, small parallel sheets are formed at the edge of the β sandwich through hydrogen bonding of strand β_{12} to β_{16} at the edge of the outer sheet, and β_1 to β_{13} at the edge of the inner sheet.

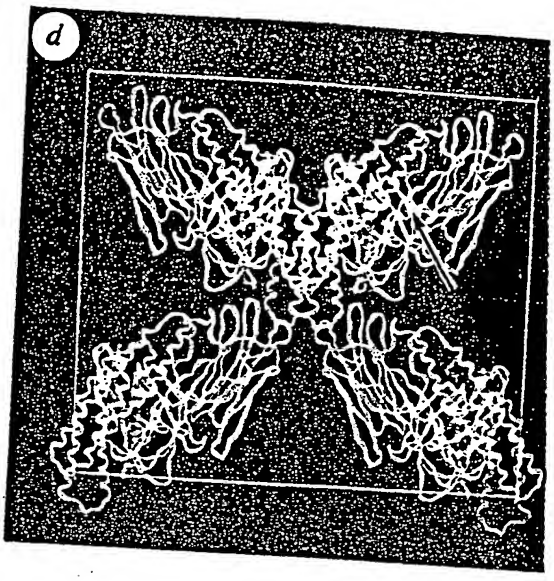
Distribution of conserved sequences. The core of the beetle toxin molecule encompassing the domain interfaces is built from the five sequence blocks that are highly conserved throughout the δ-endotoxin family[1] (Fig. 2*b*,*c*). Block 1, located in the beetle toxin sequence at residues 189–218, corresponds to the central helix (α_5) of the bundle in domain I. Block 2, residues 239–305, overlaps with the latter half of α_6, and with α_7 and β_1; the latter hydrogen-bonds to the edge of the inner sheet in domain III before forming part of the three-stranded sheet 3 in domain II. Block 3, residues 491–538, overlaps with the latter part of β_{11}, where it is hydrogen-bonded to β_1, and with the loops connecting domains II and III. The remainder of block 3 together with blocks 4 and 5, namely residues 560–569 and 633 to the C terminus, respectively, constitute the three buried strands of the inner antiparallel sheet in domain III. The high degree of conservation of internal residues implies that homologous proteins would adopt a similar fold. Using the beetle toxin structure as a model, we can therefore propose a basis for the insecticidal activity of δ-endotoxins as a family.

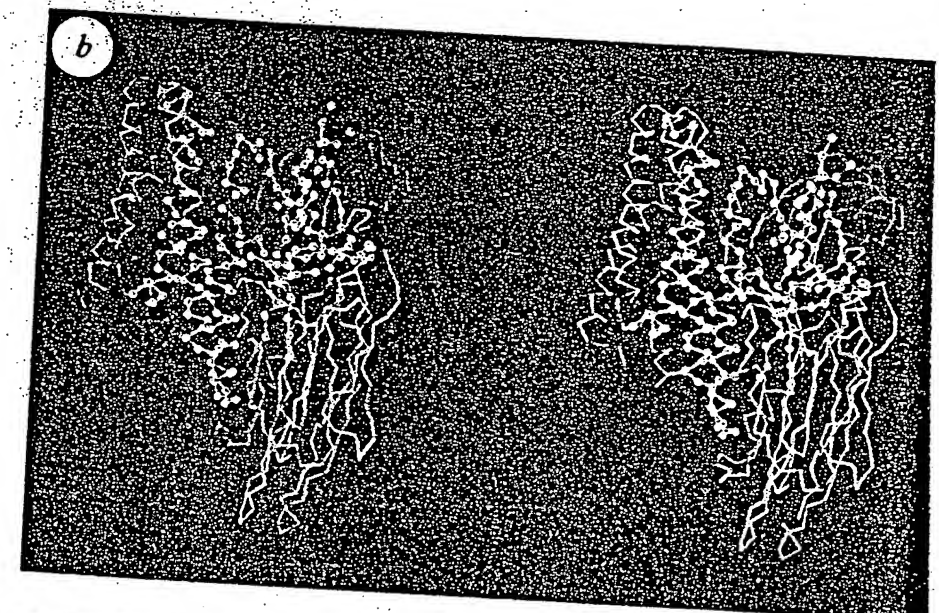
Basis of insecticidal function

Solubility. The beetle toxin crystals are isomorphous with the parasporal crystals[18,19] and show the molecular contacts responsible for solubility behaviour *in vivo*. Four intermolecular salt bridges, Asp 142–Arg 165, Asp 224–Arg 562, Asp 590–Arg 178, and Glu 223–Lys 293, are located at contacts to three different neighbouring molecules. Such salt bridges keep the protoxin crystals insoluble until exposed to the extreme pHs in the insect midgut.

Proteolytic activation. Pro-δ-endotoxins have M_rs of either ~130K or ~70K. Activation by larval gut proteases removes the C-terminal half of the larger protoxins[30,31] and cleaves them at residue 28 or 29 from the N terminus. The smaller protoxins, such as that of the beetle toxin, are processed only at the N







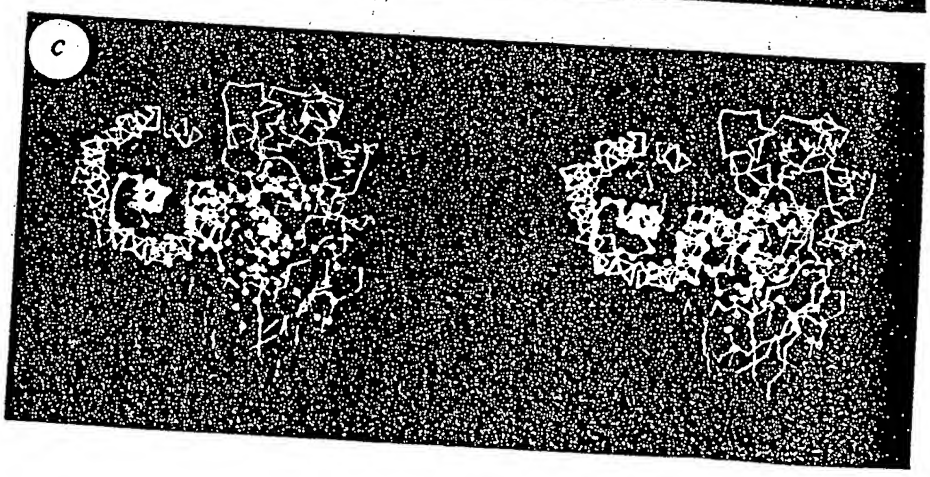


FIG. 2 Overview. *a*, Schematic ribbon representation of the beetle toxin showing the domain organization. Secondary structure assignments are given by Yasspa within program *O* (ref. 23). The polypeptide pathway is indicated by colouring the chain in the rainbow order, from red at the N terminus to blue at the C terminus. The three domains are: I, a seven-helix bundle (upper left); II, a three-sheet assembly (bottom); and III, a β sandwich (upper right). This and all following illustrations of the structure are made with the program MOLSCRIPT[47]. *b* and *c*, Cα trace (stereoview) of the molecule with the five conserved sequence blocks indicated by small beads at their Cα positions. In *b* the view is as in *a*, and in *c* it is down the central helix of the bundle from the bulky end of the molecule; *c* shows that the central helix of domain I and the inner sheet of domain III are conserved; *b* shows that the helices at the domain I–II interface and the loops at the domain II–III interface are also conserved. Note in *c* the helix packing of six around one in domain I. *d*, The solvent channel in the $C222_1$ lattice viewed along the *c* axis. One half of the unit cell thickness is shown, containing four molecules. The other half of the cell is related to this by a two-fold rotation about horizontal axes (blue lines) at $(\frac{1}{2}, y, \pm\frac{1}{4})$. The stacking of both layers leaves solvent channels that traverse the cell along the *c* direction. The N terminus of the molecule (arrow) is accessible from these channels.

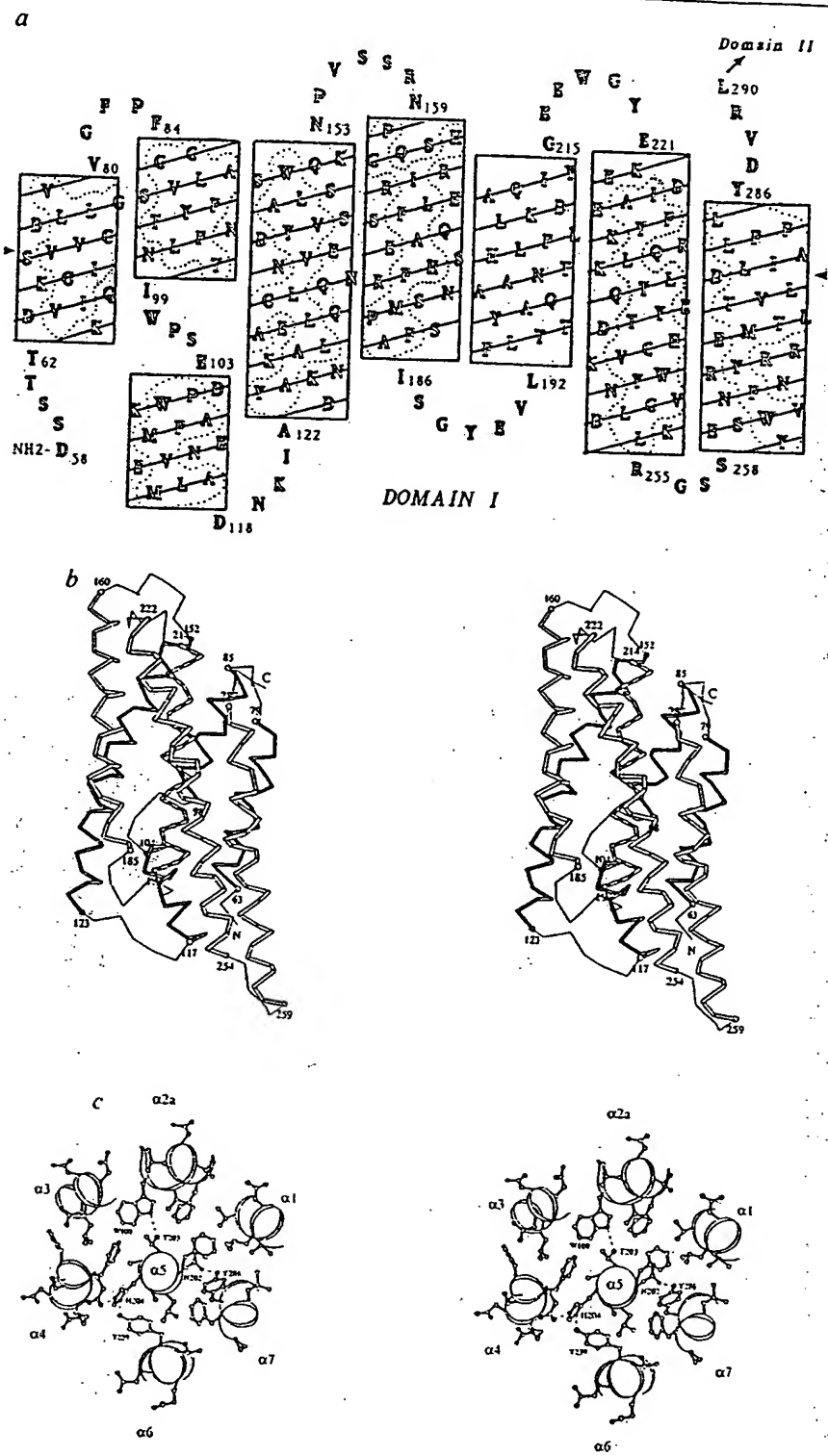


FIG. 3 The seven-helix bundle. *a*, Helical nets showing the position of amino-acid residues along the 7 helices: α_1 (63–79); α_2 (α_{2a}, 85–98 and α_{2b}, 104–117), σ_3 (123–152), α_4 (160–185), α_5 (193–214), α_6 (222–254) and α_7 (259–285). The cylindrical surface of the helices are cut longitudinally on the side facing the solvent and flattened to give a view from the interior of the bundle. The top of the drawing corresponds to the bulky end of the whole molecule. Owing to tilting of the outer helices, different helices are in register vertically only at a level indicated by two arrows pointed at α_1 and α_7; α_5 is the central helix. Dotted curves outline the strip of hydrophobic residues down the inward surface of the other six helices. *b*, Cα trace (stereoview) for the bundle viewed perpendicular to α_5. The relative tilt of the outer helices to α_5 and that between adjacent outer heleices are both about 20°. The Cα trace is shaded grey over helices α1 to α3 in the back, striped over helix α5 in the centre, and white over helices α4, α5, and α7 in the front. *c*, Cross-section of the bundle at the level indicated by the arrows in *a*, viewed from the bulky end of the molecule. The hellical backbone is represented by curly ribbons passing through the Cα positions. The outer helices are positioned roughly hexagonally around the central one and tilted relative to it, so the bundle forms a left-handed superhelix. The aromatic side chains are packed in an edge-to-face fashion. Hydrogen bonds are shown for side-chain atoms.

terminus[19,32] where about 50 residues are removed. The activated δ-endotoxins show a conserved C-terminus, so-called sequence block 5 (ref. 1). Its position as the middle strand of the buried β sheet in domain III precludes further processing from the C terminus. In fact deletion from this site by 4 to 8 residues results in inactive mutants with altered solubility and immunogenicity[30,33–35]. This is not surprising as the inner sheet can be expected to play a critical part in the structural integrity and stability of the toxins through interaction with the helical bundle.

At the N-terminal cleavage sites the different protoxin sequences show locally similar hydropathy profiles[36,37], which would be consistent with a common topology for the N-terminal region of the activated toxins as seen in the helical bundle of the beetle toxin. In crystals of the beetle toxin, the N terminus at the start of helix α_1 borders on a large solvent channel of about 30 Å diameter that crosses the unit cell along the *c* direction (Fig. 2*d*). This channel could allow access of sporulation-associated proteases to the cleavage site in parasporal crystals[19].

Receptor binding. The insecticidal selectivity of δ-endotoxins is due to high-affinity binding to specific membrane receptors[7–9,38], which in three cases seem to be glycoproteins[38–40]. For several δ-endotoxins the specificity-determining regions have been delimited by exchanging sequence segments between closely related toxins of differing specificities[13–15]. Guided by the location of secondary structures in the beetle toxin, a plausible alignment of δ-endotoxin sequences was made for the non-

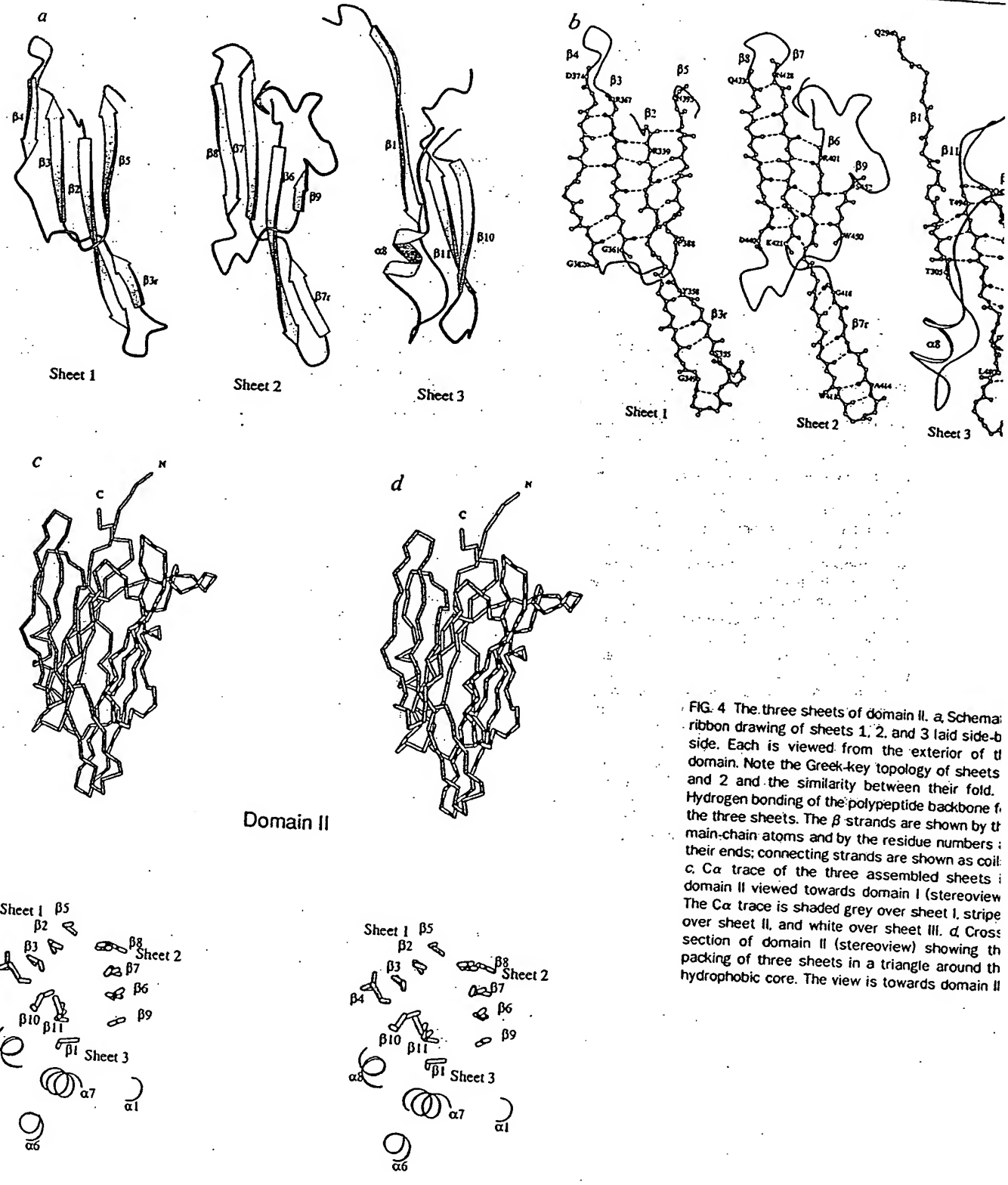
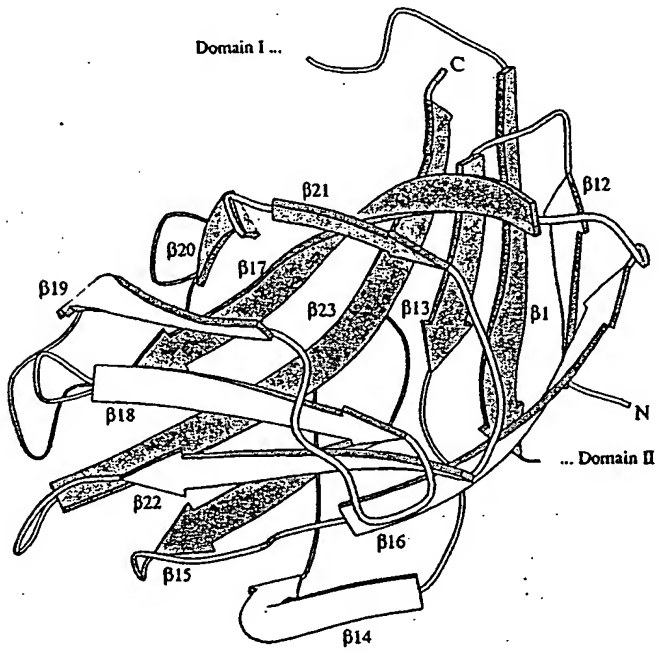


FIG. 4 The three sheets of domain II. *a*, Schema ribbon drawing of sheets 1, 2, and 3 laid side-b side. Each is viewed from the exterior of t domain. Note the Greek-key topology of sheets and 2 and the similarity between their fold. Hydrogen bonding of the polypeptide backbone f the three sheets. The β strands are shown by t main-chain atoms and by the residue numbers their ends; connecting strands are shown as coil *c*, Cα trace of the three assembled sheets domain II viewed towards domain I (stereoview The Cα trace is shaded grey over sheet I, stripe over sheet II, and white over sheet III. *d*, Cross section of domain II (stereoview) showing th packing of three sheets in a triangle around th hydrophobic core. The view is towards domain II

conserved regions (ref. 12, and T. C. Hodgman, unpublished results). Hence the genetically identified specificity-determining regions can be mapped to equivalent positions in the beetle toxin structure, and these fall mainly in domain II. For instance, the dual specificity of CryIIA for Lepidoptera and Diptera, as distinct from the Lepidoptera specificity in the closely related CryIIB, is determined by residues 307–382 of their sequences[14], which corresponds roughly to sheet 1 (Fig. 4*a*) plus strand β_6 in sheet 2 and the loop leading up to β_7, whereas the Lepidoptera specificity of CryIIB is dependent on a longer segment[14] that would include both inner strands of sheet 2. Similarly, the toxicities of CryIA(a) and CryIA(c) to two lepidopteran insects depend on three segments termed x, y and z (ref. 15): amino-acid substitutions in y can reduce toxicity by up to 2,000-fold, and segments x and y interact in determining specificity. Aligned with the beetle toxin structure, segment x corresponds roughly to the outer strands β_4 and β_5 of sheet 1 and the whole of sheet 2, including the loop entering β_{10} in sheet 3; y corresponds to



Domain III

FIG. 5 Domain III, schematic ribbon representation of the β sandwich. β strands forming the inner sheet are shaded grey. The topology of an eight-stranded 'jelly-roll' can be seen by following the β hairpin starting with β_{13}, β_{15} and β_{23} in the inner sheet, continuing to β_{16} and β_{22} in the outer sheet, then β_{17} and β_{21}, β_{20} in the inner sheet, and ending with β_{18} and β_{19} in the outer sheet. β_{14} is an excursion from the hairpin and forms a fifth antiparallel strand of the outer sheet. Small parallel β sheets are added to one edge of the β sandwich, by hydrogen bonding of β_1 to β_{13} in the inner sheet and β_{12} to β_{16} in the outer sheet. Residue numbers in the β strands are: β_{12}, 502–506; β_{13}, 509–513; β_{14}, 519–525; β_{15}, 536–541; β_{16}, 547–554; β_{17}, 558–569; β_{18}, 573–579; β_{19}, 585–591; β_{20}, 604–609; β_{21}, 611–614; β_{22}, 619–625; and β_{23}, 631–643.

strand β_{10} of sheet 3 and the loop connecting β_{10} and β_{11}; and z extends from β_{11} to the C-terminal activation site. Furthermore, the interaction between x and y can be understood in terms of the proximity between β_4 on the edge of sheet 1 and β_{10} on the edge of sheet 3. Although z was inferred[15] to extend into domain III, the combined evidence from genetics and receptor-binding assays *in vitro* for Lepidoptera toxins[9,41] correlates receptor recognition with sequence variations within domain II. We note that the β ribbons from all three sheets terminate in loops in a small region on the molecular apex, in a manner reminiscent of the complementarity-determining region of immunoglobins.

Pore formation. The common mechanism of epithelial cell disruption by δ-endotoxins of widely different specificities is believed to be the formation of lytic pores of 10 to 20 Å diameter in the insect membrane[10]. The structure of the beetle toxin displays an apparatus for pore formation in the long, hydrophobic and amphipathic helices of domain I which could penetrate the membrane. Between the crystal structure in which the bouquet-like helical bundle internalizes all the hydrophobic surfaces, and the unknown pore structure where hydrophobic surfaces would be in intimate contact with the membrane lipids, large conformation changes must occur. In the absence of a full characterization of the pore-forming process, we propose the following by extrapolation from the crystal structure.

The trigger for the conformational changes may be provided by receptor binding and the consequent interaction of toxin with the membrane bilayer. Membrane insertion follows rapidly, so that a major part of the bound δ-endotoxin cannot be displaced from the brush-border vesicles by other toxins recognizing the same receptor sites[7,9]. As domain II and probably its apical region are most likely to bind the membrane receptors, the helices are expected to insert with the 'domain II end' (see Fig. 2*a*) oriented towards the cytoplasm. If helical hairpins are to initiate the membrane penetration, as probably happens for colicin[28,42,43], they will probably be linked at the domain II end. So either of the helix pairs α_6–α_7 or α_4–α_5 could be the likely initiator. The α_6–α_7 pair is favoured because it forms part of the conserved interface with domain II and is well positioned to sense the receptor binding. On the other hand, helix α_5 is the most conserved throughout the family of δ-endotoxins. Point mutations in α_5 reduce toxicity of a Lepidoptera toxin without reducing binding to membranes[44]. Proteolysis in the interhelical loops at the domain III end, as in the α_3–α_4 loop[19,32], may facilitate release of the helix pairs from the tertiary structure of the bundle. The insertion of a hairpin can create a defect in the membrane, allowing the rest of domain I to participate in pore formation in a cooperative manner. □

Received 22 July; accepted 19 September 1991.

1. Höfte, H. & Whiteley, H. R., *Microbiol. Rev.* **53**, 242–255 (1989).
2. Ellar, D. J. *et al.* in *Molecular Biology of Microbial Differentiation* (eds Hoch, J. A. & Setlow, P.) 230–240 (Am. Soc. Microbiol., Washington, DC, 1985).
3. Wilcox, E. R. *et al.* in *Protein Engineering: Applications in Science, Medicine and Industry* (eds Inouye, M. & Sarma, R.) 395–413 (Academic, New York, 1986).
4. Vaeck, M. *et al. Nature* **328**, 33–37 (1987).
5. Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, D. L. & Fischhoff, D. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3324–3328 (1991).
6. Barton, K. A., Whiteley, H. R. & Yang, N. S. *Plant Physiol.* **85**, 1103–1109 (1987).
7. Hofmann, C., Lüthy, P., Hütter, R. & Pliska, V. *Eur. J. Biochem.* **173**, 85–91 (1988).
8. Hofmann, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7844–7848 (1988).
9. Van Rie, J., Jansens, S., Höfte, H., Degheele, D. & Van Mellaert, H. *Appl. Envir. Microbiol.* **56**, 1378–1385 (1990).
10. Knowles, B. H. & Ellar, D. J. *Biochim. biophys. Acta* **924**, 509–518 (1987).
11. Endo, Y. & Nishitsutsuji-Uwo, J. *J. Invertebr. Path.* **36**, 90–103 (1980).
12. Hodgman, T. C. & Ellar, D. J. *J. DNA Sequ. Map.* **1**, 97–106 (1990).
13. Ge, A. Z., Shivarova, N. I. & Dean, D. H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4037–4041 (1989).
14. Widner, W. R. & Whiteley, H. R. *J. Bact.* **172**, 2826–2832 (1990).
15. Schnepf, H. E., Tomczak, K., Ortega, J. P. & Whiteley, H. R. *J. biol. Chem.* **265**, 20923–20939 (1990).
16. Krieg, A., Huger, A. M., Langenbruch, G. A. & Schnetter, W. *J. appl. Entomol.* **96**, 500–508 (1983).
17. Höfte, H., Seurinck, J., Van Houtven, A. & Vaeck, M. *Nucleic Acids Res.* **15**, 7183 (1987).
18. Li, J., Henderson, R., Carroll, J. & Ellar, D. *J. molec. Biol.* **199**, 543–545 (1988).
19. Carroll, J., Li, J. & Ellar, D. *J. Biochem. J.* **261**, 99–105 (1989).
20. Arndt, U. W. *Meth. Enzym.* **114**, 472–485 (1985).
21. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
22. Leslie, A. G. W. *Acta crystallogr.* **A43**, 134–136 (1987).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, N. *Acta crystallogr.* **A47**, 110–119 (1991).
24. Brünger, A. T. *J. molec. Biol.* **203**, 803–816 (1988).
25. Chothia, C. & Finkelstein, A. V. *A. Rev. Biochem.* **59**, 1007–1039 (1990).
26. Chothia, C. *J. molec. Biol.* **105**, 1–15 (1976).
27. Burley, S. K. & Petsko, G. A. *Science* **229**, 23–28 (1985).
28. Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. *Nature* **337**, 93–96 (1989).
29. Richardson, J. S. *Adv. Prot. Chem.* **34**, 167–339 (1981).
30. Höfte, H. *et al. Eur. J. Biochem.* **161**, 273–280 (1986).
31. Choma, C. T. & Kaplan, H. *Biochemistry* **29**, 10971–10977 (1990).
32. Nicholls, C. N., Ahmad, W. & Ellar, D. J. *J. Bact.* **171**, 5141–5147 (1989).
33. Macintosh, S. C., McPherson, S. L., Perlak, F. J., Marrone, P. G. & Fuchs, R. L. *Biochem. biophys. Res. Commun.* **170**, 665–672 (1990).
34. Schnepf, H. E. & Whiteley, H. R. *J. biol. Chem.* **260**, 6273–6280 (1985).
35. Adang, M. J. *et al. Gene* **36**, 289–300 (1985).
36. Sekar, R., Thompson, D. V., Maroney, M. J., Bookland, R. G. & Adang, M. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7036–7040 (1987).
37. Chungjatupornchai, W., Höfte, H., Seurinck, J., Angsuthanasombat, C. & Vaeck, M. *Eur. J. Biochem.* **173**, 9–16 (1988).
38. Knowles, B. H., Thomas, W. E. & Ellar, D. J. *FEBS Lett.* **168**, 197–202 (1984).
39. Knowles, B. H. & Ellar, D. J. *J. Cell Sci.* **83**, 89–101 (1986).
40. Haider, M. Z. & Ellar, D. J. *Biochem. J.* **248**, 197–201 (1987).
41. Visser, B., Munsterman, E., Stoker, A. & Dirske, W. G. *J. Bact.* **172**, 6783–6788 (1990).
42. Lakey, J. H., Baty, D. & Pattus, F. *J. molec. Biol.* **218**, 639–653 (1991).
43. Song, H. Y., Cohen, F. S. & Cramer, W. A. *J. Bact.* **173**, 2927–2934 (1991).
44. Ahmad, W. & Ellar, D. J. *FEMS Microbiol. Lett.* **68**, 97–104 (1990).
45. Messerschmidt, A. & Pflugrath, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
46. Read, R. J. *Acta crystallogr.* **A42**, 140–149 (1986).
47. Krautis, P. *J. appl. Crystallogr.* (in the press).

ACKNOWLEDGEMENTS. We thank P. R. Evans, A. G. W. Leslie and R. Henderson for advice and encouragement; K. Wilson and Z. Dauter for help with the image plate system; K. Nagai and P. J. McLaughlin for help in collecting film data; T. A. Jones for advice on model building; SERC Daresbury Laboratory and EMBL Outstation at DESY for use of synchrotron facilities; and T. Woollard and K. Hopkins for maintaining the rotating anodes. D.J.E. and J.C. acknowledge the support of the AFRC.